Index Numbers

WILEY PUBLICATIONS IN STATISTICS

Walter A. Shewhart, Editor

Mathematical Statistics

DWYER—Linear Computations.

FISHER—Contributions to Mathematical Statistics.

WALD—Statistical Decision Functions.

FELLER—An Introduction to Probability Theory and Its Applications, Volume One.

WALD—Sequential Analysis.

HOEL—Introduction to Mathematical Statistics.

Applied Statistics

MUDGETT—Index Numbers.

TIPPETT—Technological Applications of Statistics.

DEMING—Some Theory of Sampling.

COCHRAN and COX—Experimental Designs.

RICE—Control Charts.

DODGE and ROMIG—Sampling Inspection Tables.

Related Books of Interest to Statisticians

HAUSER and LEONARD—Government Statistics for Business Use.

Index Numbers

BRUCE D. MUDGETT
Professor of Economics and Statistics
University of Minnesota

New York · John Wiley & Sons, Inc.
London · Chapman & Hall, Limited

PRINTED IN THE UNITED STATES OF AMERICA

Preface

INDEX NUMBERS ARE TOOLS OF ECONOMIC MEASUREMENT AND ARE already venerable. One of the earliest puzzles that occupied men's minds in the dawning days of modern industrialism was what made prices increase. We know now that silver poured into Europe from the Americas for several centuries after Columbus' famous voyage, and we also know that by the mid-eighteenth century there was much concern over the great increase in prices on the European continent. The question how much prices had increased demanded an answer. As the pages of this study will tell, an Italian, Carli, provided the first answer. When he found prices of three commodities at two dates separated by 250 years and then took an average of the three price relatives, he had a measurement of the percentage change in price levels from the year 1500 to the year 1750. This is probably the first index number on record in the modern sense of the term.

Interestingly enough, progress in both the theory and the practice of index numbers has seemed always to be related to periods of rapid upheaval in price structures. In the history of Europe and North America since 1492, several such radical price movements can be traced. There were the silver imports into Europe referred to above; there were gold imports into Europe from California and Australia after 1848; there were the long-continued decline in prices in both Europe and America from 1865 to 1895 and then the gradual increase in price levels after 1895, culminating in the worldwide inflation of World War I and followed by post-war deflation; and, still later, there has been the price inflation of World War II.

The fact of price change brought the need for measurement; and, in association with each of the price upheavals listed above, developments occurred in the theory of index numbers. Of modern writers, the first was Jevons, who sought a measurement of the effects of gold imports on price levels after 1848. Marshall and Edgeworth wrote in the 1880's, when the problem of bimetalism and its relation to price levels was a political issue of great importance in both England and the United States; Mitchell wrote for the United States Bureau of Labor Statistics when the accuracy of its measurement of the whole-sale price level was brought into question at the end of the first decade

of this century. Irving Fisher's book of 1923 was an avowed effort to establish standards of accurate measurement and to evaluate results.

Through all these times, the development of theory and method has moved in a direct way toward better measurement. To illustrate only by the growth of practice in the United States, one of our first important index numbers, the wholesale price index of the U. S. Bureau of Labor Statistics, was established on a permanent basis in 1902 and was then a simple unweighted average of relatives. Its accuracy was questioned before many years, and the bureau then sought the assistance of Wesley Mitchell in a revision. The Mitchell bulletin of 1915 established the principle that weighting was necessary, and its most important single effect was the introduction of the fixed-weight aggregative formula. This formula is almost universally used today in the United States.

The end of World War I brought widespread discussion of price levels and extended arguments over index number measurements. Fisher's book in 1923 probably represents the peak of this development, which was characterized by a feature radically at variance with current practice: an emphasis upon the need for using current weights. This method contrasts with that of fixed weights (to be revised maybe once in ten years), which had been recommended in the Mitchell bulletin and which had by then become general practice. Though the critics of Fisher's book were most severe in some of their condemnations, it is significant that the practice of the U. S. Bureau of Labor Statistics was changed after 1923 with especial reference to weights: for several years they actually changed weights every two years, as new data became available.

Many students of index numbers may be unwilling to accept the view that the price upheaval of World War II and the long argument after 1943 about the accuracy of the consumers' price index were the most important factors in modern times in pointing up the weakness of the fixed-weight aggregative formula for index numbers. It is my view, however, that fixed weights are the greatest weakness of modern indexes, and one purpose of the following pages is to show that a basic analysis of the problem of index number measurements leads infallibly to this conclusion.

It must be emphasized, finally, that I have not dealt in any way with another, and a vitally important, part of the problem of index number measurement, namely, the accuracy of the original price and quantity data. No book has been written on this subject as yet, although much progress has been made by government bureaus that

specialize in the construction of index numbers of national scope. The subject deserves a book written by specialists in this field, and I hope that some of the able technicians now working in government offices on this problem can be persuaded to write such a book. The following pages will never compete with it; I hope they may be a complement to it.

BRUCE D. MUDGETT

University of Minnesota
April 25, 1951

Contents

PART I · THE MEASUREMENT OF PRICE AND QUANTITY CHANGE

Part I • The Measurement of Price and Quantity Change

CHAPTER 1

Introduction

1.1. WIDE USE OF INDEX NUMBERS

Price and quantity indexes have in recent years become tools of daily use in the interpretation of current economic conditions. Changes in price levels are watched eagerly by employers and by laborers. Wages have been tied to cost-of-living indexes many years in England. The same has been true in the United States in recent years, and the practice has been increasing. The index of industrial production of the Federal Reserve Board is published a few days after the close of each month and is awaited eagerly by a national audience of businessmen and students of the national economy for the "information" it furnishes on the current position of national production—basically, productive activity in the mining and manufacturing fields. Farmers in the United States, with equal eagerness, watch the index of parity prices, for there is now a national law that requires government support of certain farm prices in order to maintain a specified parity relationship with a farm price situation that held many years ago—one of those comparison bases that is presumed to represent *normal* conditions, or at least conditions more generally acceptable than others to those groups whose influence brought about the passage of the parity law.

1.2. MEASUREMENT DISTINGUISHED FROM POLICY

The use of index numbers to measure certain magnitudes in an economy must be rigorously separated from any justification of particular economic policies which their use is intended to implement. If a businessman decides to tie his wage policy to living costs, an accurate index of living costs is a basic element in carrying out such a policy. Or, if Congress passes a law requiring that farm prices shall be maintained on a parity relationship with other prices such as existed at some earlier time, it becomes an essential feature of the operation of the law that the parity relationship be measured, and this measurement produces an index number of prices. Of course, changes in the direction

of policy will occur from time to time both with respect to firms and individuals and with respect to governments. But there will probably be no lessening in the demand for measurements of various elements of our economy; indeed, every indication is that there will be greatly increased demands for measurements of economic magnitudes. Among these demands it is to be expected that index numbers of prices and quantities will be prominent items.

1.3. NEED FOR MORE REFINED MEASUREMENTS

It is therefore not amiss for the technicians to take stock of their tools from time to time and to consider whether the tools can be refined and made more efficient. It has been said that the greatest advances in science come only under the pressure of need, and examples can be cited in economics to give some support to such a generalization. Indeed, the history of index numbers would seem to bear it out. Particularly today, with many of the most important indexes constructed by government bureaus, severe criticism may be needed to force change in *established* and *accepted* methods.

Price index numbers originated[1] in 1764 when Carli attempted to measure changes in the purchasing power of money consequent upon the importation of silver from America. Prices had gone up; the question was: How much? Carli supplied the world with the first price index number formula, a simple arithmetic average of price relatives. The United States Bureau of Labor Statistics used this formula for measuring wholesale prices in the United States in the early years of the current century, and the shift in 1915 to a different and better formula came only after much criticism of the bureau's index in those early years. In the course of World War II many objections were raised to the measurement of industrial production in the index of the Federal Reserve Board, and these criticisms must have played their minor part in the revisions of the early 1940's, along with the undoubted desire of the board's technical staff to improve the index whenever it was possible. Indeed, such criticisms may often be the fillip needed to stir bureau directors or Congress or the American people to increase their financial support of this sort of scientific work. There was an extended argument between the AFL-CIO labor representatives and the U. S. Bureau of Labor Statistics over the cost-of-living index of the bureau, and although this argument did not bring

[1] Dutot in 1738 had compared sums of prices at two periods. It could be claimed from this fact that first credit for the idea of index numbers belongs to Dutot. But see Mitchell, Bull. 284, p. 7.

about an immediate revision of the index it did produce an extended investigation of the criticisms by an imposing committee of experts.

1.4. TENTATIVE JUDGMENTS ON CURRENT METHODS

Present methods of constructing index numbers of price and quantity can be very inaccurate. The charge that the federal bureau's cost-of-living index failed by a considerable percentage to measure the true change in the cost of living may well be true; and inaccuracies of this sort are likely to show up in times of exceptional economic change such as occurred during World War II. It is further believed that some of the tasks which index numbers are expected to perform can be done very effectively, such as the measurement of year-to-year changes in price levels, whereas sometimes index numbers may produce measured results that have no counterpart in reality and that should be discarded. An example of the latter results is the measurement of price level changes from colonial times to the present. It is the purpose of this study to give concrete embodiment to these general notions. It is too much to expect that there will be immediate and general acceptance of all the ideas developed in the pages to follow, but it can at least be hoped that they will receive the serious consideration and criticism that the author believes they deserve from competent students of the subject.

CHAPTER 2

The Meaning of Price and Quantity Change

2.1. SOME REAL PROBLEMS OF A CHANGING WORLD

Consider the following topics, each of which has at some time, perhaps often, been a matter of serious study in the economies of modern nations: (1) the purchasing power of money, (2) inflation or deflation, (3) growth of or change in real-goods production, (4) total value production, or consumption (or the measurement of the national income).

As stated above, one of the first index numbers of which we know was constructed in 1764 by Carli, who was interested in discovering the effect, upon the purchasing power of money, of silver imports into Europe after the discovery of America. Carli had only three commodities, grain, wine, and oil, available to compare prices at two dates separated by 250 years; but, such as it was, he got an answer to his question about purchasing power by calculating the percentage change in price (price relative) for each of the three commodities between the years 1500 and 1750 and then taking a simple average of the three ratios. This was his measurement of the price change that had occurred. We would now say that its reciprocal measured the change in the purchasing power of money, over this 250-year interval. Characteristically, it was this problem of money purchasing power that first occupied students of index numbers and that is responsible for the main body of index number literature down to practically modern times—at least until the beginning of World War I. Jevons' study, "A Serious Fall in the Value of Gold Ascertained" (1863), falls in this category. One of Alfred Marshall's early manuscripts (*Contemporary Review*, 1887) dealt with the problem and with an index number solution. Irving Fisher's *The Purchasing Power of Money*, published in 1911, points clearly to the overwhelming emphasis, up to this date, on the purchasing-power concept, as does Walsh's *The Measurement of General Exchange Value* (1901).

It is not, of course, implied that inflation and deflation are basically different from purchasing power of money. They might be said to represent the other face of the coin. Inflation means high prices, and they, in turn, mean low purchasing power. But the index number literature since about 1914 has used the terms inflation and deflation much more than the term purchasing power. The reason for this change in emphasis is that in recent times we have been more interested in inflation applied to some partial segment of the whole economy, whereas the literature of "purchasing power" was concerned with general purchasing power or with the problem of the relationship of all money for exchange and all goods seeking exchange. Nevertheless, the thread of a common idea runs through all our index number literature—the notion of a price level, or a plateau of prices, applying to the total or to some segment of the economy. The problem is to measure the difference between the height of the plateau at one period of time and that at another. If a solution is found, it will be called an index number.

The origin of the idea of changes in the quantity of real-goods production (or consumption) doubtless goes back to the beginning of thinking about the theory of value, but the thought of measuring changes in production in quantitative, rather than in value, terms probably came much later. It is difficult to say just when the idea of a quantity index first appeared in the literature; but it is a fact, nevertheless, that it did not play an important part in index number literature until after the beginning of the twentieth century. As implied earlier, long-continued elements of instability in prices brought discussions that produced the literature of index numbers. The writings of Jevons came at the end of a period of rising prices; Edgeworth and Marshall wrote during the long price decline from the 1860's to the 1890's. After the nineties a period of increasing prices set in, and during this period index number literature began to accumulate in great volume. Furthermore, from this time on, more and more attention was paid to index numbers of limited scope, for example, regional price indexes, wholesale, retail, cost-of-living indexes, and the like.

Indexes of production are an item in this specialization. The outstanding literature of the period, which provides the source of the above statements, includes, first, Mitchell's study for the U. S. Bureau of Labor Statistics (*Wholesale Prices*, Bull. 173, 1915, revised as Bull. 284 in 1921), which definitely pointed to the importance of specialized indexes; Pigou's *Wealth and Welfare* (1911); a huge volume of periodical literature after World War I, especially in the United States and England; and finally the publication of Fisher's *The Making of Index*

Numbers (1923). Throughout the last part of this period, say from 1910 to 1925, the literature discussed how to measure changes in real-goods production, as well as price changes. A new problem arose at this time, the measurement of the national dividend or national income. The problem is discussed at length by Pigou in his *Economics of Welfare.* The National Bureau of Economic Research, organized in New York in 1920 for pure research, set this topic as one of its earliest extensive studies and published its first volume on the subject in 1923.

It is therefore not too wide of the mark to say that index number literature is an answer to the demand for particular kinds of measurement; that this demand is great when great price change has produced instability in the economic system; and that, in the history of index numbers, problem and solution have gone hand in hand.

2.2. THESE PROBLEMS GENERALIZED

This brief reference to certain economic problems the solutions of which have created index numbers was designed to introduce some real-life examples that have concrete meaning for everybody. The purchasing-power concept may escape many persons; but few have escaped the inflations of the later 1940's, and everyone who deals with a family budget knows what the word inflation means. Stretching a limited family income over the necessities of decent living at a time when prices are advancing much faster than family incomes threatens living standards. The term inflation has to do, of course, with a set or group of prices conceived in their entirety, not with individual prices. For some prices might have gone down although it might appear to the householder that *prices* have gone up. We now accept such terms as national income (a value or dollars concept) and real income (a physical-goods concept, from which the element of price variability is absent) as meaningful expressions applicable to the economic life of nations or peoples. If, then, there is such a thing as inflation, there should be some way more or less exact to measure it. And, if American workmen today enjoy higher standards of consumption than formerly (that is, more real goods, not merely more dollars' worth), there should be a means of measuring this *more* in some quantitative sense, some way, that is, of answering the question: How much more?

When one approaches the task of measuring one of these quantities, it will be seen that certain features are always present:

1. First, there is always a comparison of two situations. The literature says that there is first the problem of making a *binary*

comparison; then follows the comparison in series. The two situations in the binary comparison may be two time periods, such as 1940 and 1941, or they may be two places, such as Detroit and Warsaw.[1] We say that the comparison may be in time or in space, but, actually, most index numbers do involve comparison in time.

2. The basic material of the two-period comparison is two sets of data (in theory completely known, in practice often very difficult to secure) referring to all the *economic goods* in each period—the facts of *actual* quantities and prices of each commodity in each period. This means that there exist prices, p, and quantities, q, of two sets of commodities; whether we can obtain the complete record is another matter. From these facts simple multiplication (of price by quantity) followed by simple addition (of such values figured for each commodity) produces two sets of total monetary values, one for each situation in the comparison. The ratio, V, of these two monetary aggregates is a *value* ratio, and in general it will have a magnitude other than unity. It will be referred to frequently as a value change, since in real life it is never equal to unity and since what follows would be in no way affected by such a circumstance anyway.

3. Finally a generalization about the components of this ratio is formulated. Since the detailed components are prices and quantities it becomes conceptually possible to formulate an idea of *total price* influence and *total quantity* influence and then to say that they jointly produce the value ratio. Herein lies the index number problem with which this book is concerned. If we use the terminology

V = value ratio
P = a measure of the total price influence in V
Q = a measure of the total quantity influence in V

the relationship of these three elements becomes

$$V = P \cdot Q$$

This is an analytical statement about the facts of historical record, namely, the p's, q's, and V, and it involves a concept of the total

[1] In the late 1920's Henry Ford wanted to fix wages in his European offices to provide the same standard of consumption as enjoyed by his employees in Detroit. Study of the technical problem brought a report by the U. S. Bureau of Labor Statistics (see *Monthly Labor Rev.*, June 1930) and a number of studies originating in the International Labour Office in Geneva.

price influence P and the total quantity influence Q that arises out of these facts.

It must be stated, finally, with the greatest of emphasis that for any index number of prices or quantities that has relevance to a real-life change from one specified situation to another (that is, to any of the problems referred to above, such as purchasing power, inflation, deflation, changing production, and consumption) one must obtain the objective facts for measurement of this change from the prices and quantities of the actual goods that are involved. If in any such comparison one goes outside these two sets of price and quantity data to seek a solution to the price or quantity index problem, it must be recognized at once that he is dealing with approximation *and either the approximation must be justified or its accuracy taken under consideration.* This point of view will recur frequently in the succeeding pages.

The measurement tool that is needed in this situation is one applying, not to single commodities, but to groups of commodities. The result in one case is a price index and in the other a quantity index, and it will, if effective, separate the value change of a specified set of commodities (the ratio above) into two parts or elements, one a measurement of the overall influence of changing prices, the other a measurement of the overall influence of changing quantities. The two elements together explain the total change in value. The next problem is how to construct this tool.

CHAPTER 3

The Problem of Measurement

3.1. A SPECIFIED FIELD OF COMMODITY STUDY

In the first place an index number of price or quantity must of necessity refer to an enumerable set of commodities. This is doubtless what some of the literature refers to when it states that the purpose of an index number must first be clearly defined Thus Mitchell in his study for the U. S. Bureau of Labor Statistics [1] speaks of a general-purpose index and of special-purpose indexes. Mitchell has been criticized for this statement, especially by Fisher, who says that the purpose of an index is fixed once the set of commodities has been decided upon. The argument of the previous chapter is of course in agreement with Fisher's position. If the task of an index number is to measure that part of a historical change in value which is due to price change (or quantity change), then *purpose* can mean nothing except as it applies to a *specified* set of commodities. Let us consider possible examples. If Fisher's equation of exchange, $P \cdot T = MV + M'V'$, is concerned with establishing a relationship between all money of exchange and all goods for exchange,[2] then the money values and the quantities of goods represent the raw materials out of which index numbers are to be constructed. So the purpose here is the completely general one of measuring the change in the level of all prices. And with equal emphasis it must be said that this index number can have no narrower purpose (such, for example, as the measurement of wholesale prices) so long as we are talking of a precise measurement and not of an approximation. Of course, all prices in an economy are interrelated and there are high correlations between certain groups, and so in the absence of exact measures in a defined field it has often been necessary to accept approximations from a related field. So it was that before we had special

[1] Bull. 284, pp. 23–25.

[2] In this form of the equation of exchange, P and T are symbols, respectively, for price level and for quantity level or trade; M and M' refer to money and money substitutes, and V and V' to their respective velocities of circulation.

farm price index numbers the special subindex for wholesale prices of farm products was used as a substitute. That this substitute is no longer needed since the specific indexes are now available is clear evidence that the students of farm prices considered the substitute to be inadequate.

All indexes today are therefore special indexes in the sense in which the term is here used; that is, there is a specific list of commodities involved for which prices (or quantities) have been measured. Thus the U. S. Bureau of Labor Statistics wholesale price index applies to all wholesale prices in the United States as the term wholesale has been defined by the bureau. The U. S. Bureau of Labor Statistics cost-of-living index (new name: index of consumer prices) applies only to a particular income group (wage earners and clerical workers) in a special list of thirty-four large cities in the United States. The Federal Reserve Board index of industrial production covers manufacturing (fifty-three series) and mining (eighteen series), and it is therefore not a measure of overall production because of the significant fields of production omitted, for example, wholesale and retail trade, construction industry, and agriculture. There are industry indexes (the fertilizer price index), regional indexes (many state farm price indexes), etc. The list could be extended by many items before being in any sense complete. And completeness would be almost unattainable, since many firms construct price and sometimes quantity indexes for their own use and do not make them available to the public.

3.2. DATA: HISTORICAL RECORDS OF PRICES AND QUANTITIES

Once the set of commodities for which index numbers are to be constructed has been decided upon, it must be recognized that a set of prices and of quantities of these commodities exists for each and every situation that is to be compared. The p-q price-quantity notation and the index number notation that have been practically standard in the literature are, with some additions, as given herewith:

Prices: $p', p'', \cdots, p^{(N)}$ for N commodities.
Quantities: $q', q'', \cdots, q^{(N)}$ for N commodities.
Values: $v' = p'q'$, etc.
Subscripts: 0, 1, 2, $\cdots$, k refer to the several situations to be compared. They may be consecutive or not, as required in the context, and will be referred to hereafter as *years*. Thus: year 0 equals the base year, year 1 equals the given year 1, $\cdots$, year k equals the given year k. Then p_0 refers to a base-year price,

p_1 to a given-year price, etc.; similarly for quantities and values. In some instances it will be convenient to refer to 1 as the base year, 0 as the given year. This will cause no confusion, as the context will make the meaning clear.

For index numbers:

In general: I_{01} means (some) index for the year 1 on the 0 year as base without specification as to whether it is a price or a quantity index.

For prices: P_{01}.

For quantities: Q_{01}.

(Additional notation will appear when needed.)

Now it is necessary to recognize that a defined set or group of commodities is not always a constant list and that this inconstancy is almost universally met with in practice. We are always comparing at least two situations (two years); the set of all foods at retail, for example, will include one list, say N_0, in the base period but another list, N_1, in the given period, and many commodities may be present in both years if the years are consecutive, but not so many if the two years are far apart. Thus, if one listed all foods at retail in Boston markets for the years 1947 and 1948, most items would be found in both yearly lists, but there would be a few exceptions. On the other hand, if the two years in question were 1776 and 1946, the proportion of items common to the two lists would undoubtedly be smaller. The greater the field of commodities involved, the greater the discrepancy is likely to be. This discrepancy has importance for index number measurement and therefore must be stated with precision and in symbolic notation:

Year	Prices	Quantities
0	$p_0', p_0'', \cdots, p_0^{(N_0)}$	$q_0', q_0'', \cdots, q_0^{(N_0)}$
1	$p_1', p_1'', \cdots, p_1^{(N_1)}$	$q_1', q_1'', \cdots, q_1^{(N_1)}$
.	.	.
.	.	.
.	.	.
k	$p_k', p_k'', \cdots, p_k^{(N_k)}$	$q_k', q_k'', \cdots, q_k^{(N_k)}$

From this basic and *complete* set of data, since by hypothesis it represents all the prices and quantities of all commodities existing (or involved) in these several years, it is possible to obtain several value totals with which we will be concerned in index number construction. For example, there is a *complete value total* for each year, obtained by multiplying each quantity by its price and adding all such items. A second total may be found for a given set of commodities (N_{01}) that are found in *both* of the years 0 and 1; and a third and still smaller

total (n_{01}) items may be used as a sample to represent the whole set, N_{01}. This third set is almost universally the one with which the problem of price and quantity measurement is concerned, for in practice it is almost never possible to obtain complete information on prices and/or quantities of *all* commodities in any defined group. In symbols the above statement would be:

1. Total values for periods 0, 1, 2, etc., are

$$\sum^{N_0} p_0q_0; \qquad \sum^{N_1} p_1q_1; \qquad \sum^{N_2} p_2q_2; \qquad \text{etc.}$$

2. Values of all goods common to two periods, 0, 1, are

$$\sum^{N_{01}} p_0q_0; \qquad \sum^{N_{01}} p_1q_1$$

3. Values for a sample of n goods common to two periods, 0, 1, are

$$\sum^{n_{01}} p_0q_0; \qquad \sum^{n_{01}} p_1q_1$$

It will be well to introduce special terms to emphasize these distinctions. Thus, N_0, N_1, N_2, etc., refer to *all commodities* of the several periods 0, 1, 2, etc. N_{01} and n_{01} refer to commodities found in both of the two specified periods 0 and 1; and, in conformity with a term that has been common to the literature for a long time, these commodities will be called *binary commodities.* A third term is needed to designate those commodities found in either of the two lists of a binary comparison but not in both. They will be called *unique commodities.*

Thus for comparison of periods 0 and 1 we have:

1. All commodities include N_0 and N_1.
2. All binary commodities include only N_{01}.
3. The sample of binary commodities include n_{01}.
4. All unique commodities include

$$(N_0 - N_{01}) + (N_1 - N_{01}) = N_0 + N_1 - 2N_{01}$$

It is clear from the above that neither binary nor unique commodities may be specified for a given period until a second period has been named for comparison with it.

3.3. STATEMENT OF THE PROBLEM

From the above information we can at once calculate three types of ratios which will be useful in later work. They will be referred to as $V_{01}(T)$, $V_{01}(N)$, and $V_{01}(n)$. When V_{01} is not followed by a symbol in parentheses it will generally refer to the last of the three ratios.

Each of these ratios is a comparison of two value aggregates for the periods 0 and 1, respectively; the comparisons could be for periods 0 and k or 0 and i where $i = 1, 2, \cdots, k$. The symbol $V_{01}(T)$ is the ratio of the complete sum of values in one period to the corresponding complete sum in a second period, that is, the total N_0 commodities in period 0 and the total N_1 commodities in period 1; $V_{01}(N)$ is a ratio of two sums of values, involving all those commodities that are *common* to the two periods. Thus, if the two periods are 1776 and 1946, flintlock muskets are included in 1776 and radios are included in 1946 for the ratio $V_{01}(T)$; but for $V_{01}(N)$ neither are included since the muskets were not in the market in 1946 and radios were not there in 1776. But both $V_{01}(T)$ and $V_{01}(N)$ include potatoes, which were present in the markets of both years. Furthermore, $V_{01}(n)$ will include total values for potatoes in both numerator and denominator but may well not include values for, say, paprika since, although paprika existed in both periods, either it is unimportant enough to the total to be omitted with reasonable safety or its situation as a spice may be sufficiently represented by some other spice. More detail on this matter of sample representation of a subgroup will appear later. So we have the historical fact of value change from one specified period to another, and we can define it (a) with respect to all goods in both periods, (b) with respect to all goods common to both periods, or (c) with respect to a sample of the b list.

The next step in the statement of the problem is to set up an explanation of the change indicated by any one of these value ratios. Any value ratio measures the difference (in a ratio sense) between two sums or aggregates of value, such as Σp_1q_1 and Σp_0q_0; and the standard explanation is that this value change occurs in part as a result of price change and in part as a result of quantity change. So we conceive of an overall measure of the price element and call it P_{01}; a similar measure of the quantity influence in any V_{01} is given the symbol Q_{01}. Finally, since all influences operating on the value ratio are comprehended within these two terms, *price and quantity change*,[1] we define the index number problem as the problem of breaking up a value ratio into its basic components—factors P_{01} and Q_{01}.

The argument of this section thus reduces to the following statement of the problem:

1. The fact of value change:

   $V_{01}(T)$ with respect to all goods in two periods.

[1] Remember that the terms here refer to a group phenomenon and not to individual commodities.

$V_{01}(N)$ with respect to all goods common to both periods.

$V_{01}(n)$ with respect to a sample of the *common* goods in $V_{01}(N)$.

2. The hypothesis:

   P_{01}, a factor of change due to prices.

   Q_{01}, a factor of change due to quantities.

3. The definition of the problem of index numbers:

   $V_{01} = P_{01} \cdot Q_{01}$.

(Here the value ratio may be any one of the three above.)

CHAPTER 4

Price and Quantity Variation

4.1. DEALING WITH FACTS OF HISTORY

With the acceptance of the proposition that the index number problem is to be solved by separating the value ratio into its two components P_{01} and Q_{01}, an important first step in the solution has been taken. Whatever the form in detail which the solution may take, *it is irretrievably tied to the historical record.* Index numbers as here defined are measures related to the purchasing power of money or to inflation and deflation, and inflation and deflation are matters of historical change The tool that measures them, therefore, should be fashioned to deal with the historical material, and its ultimate efficiency as an instrument of measurement should be judged by reference to the actual record of prices, quantities, and values with which men have been faced.[1] As a first step in fashioning the tool with which we are going to be concerned, it is important to review some significant facts about the variations of prices and of quantities which have come to light as a result of the work of students of this problem in the past. For it is true that these facts, or some of them, have had great influence on the development of modern notions about how to construct index numbers of price and quantity. Remembering that the index number problem will be solved if and when a measure is obtained which isolates *the price influence or the quantity influence* from the value change in a defined commodity group, we see that the immediate problem then is how to construct such a measure.

[1] This statement is intended in no way to deny the validity of the common practice of modern statistical theory of setting up mathematical (i.e., conceptual) models and of interpreting factual observations by their conformity to these models. Of course, since such models serve only for so long as they correctly interpret observations they must be replaced or revised when their failure to interpret has been ascertained. $P_{01} \cdot Q_{01} = V_{01}$ as here defined may be looked upon as a mathematical model with which to interpret concrete reality.

4.2. PRICE AND QUANTITY VARIATION FOR INDIVIDUAL COMMODITIES

If we are to keep close to our historical materials we must recognize that there is no group effect, until there are individual effects; that is, a price influence for the group must be in some way a combining or summing of individual price effects.

With p_1, p_0, q_1, and q_0 representing the prices and quantities of a given commodity in the periods 0 and 1, any variation in price or quantity for this commodity may be defined in either *actual* or relative terms. Define:

(*a*) Actual changes

$$p_{01} = p_1 - p_0$$

$$q_{01} = q_1 - q_0$$

(*b*) Relative changes

$$\frac{p_1}{p_0} \text{ and } \frac{q_1}{q_0}$$

The p_{01} and q_{01} measure the actual changes in price and quantity for the commodity, and these magnitudes are expressed in the same unit in which price or quantity is stated, whereas p_1/p_0 and q_1/q_0 are magnitudes in ratio or percentage terms or, as they are usually called, relatives. Thus there are two possible ways of measuring price (quantity) influence for individual commodities. Now how shall the same influence be measured for the whole group?

4.3. MEASURING THE TOTAL PRICE AND QUANTITY EFFECTS BY SUMMING ACTUAL PRICES AND QUANTITIES

1. Difficulties in Summing Actual Prices

Consider the price influence first. The problem is to find a way of combining or totaling the individual price influences or elements in order to obtain an overall measurement of what has happened to the group prices. In terms of an earlier discussion we wish to know whether the price plateau in period 1 is at the same level as that of period 0. It will be recognized at once that there are difficulties in dealing with actual prices. One may perform the arithmetical operation of summing actual variations in prices, that is, Σp_{01}, but this is a seriously defective procedure for measuring the total price influence. One single fact is sufficient to condemn it, namely, that the result is affected by the units in which the prices are quoted and therefore that

different answers can be obtained merely by changing the unit in which the price is quoted. Thus, for example, a summation including the price of steel is one figure if steel is given in price per ton and quite a different figure if steel is quoted at price per hundredweight or per pound. Consider a simple and somewhat absurd example:

Commodity	Unit	Prices (dollars)		Unit	Prices (dollars)	
		0	1		0	1
Steel	Ton	40.00	50.00	Pound	0.02	0.025
Darning needles	Package	0.25	0.40	Gross of packages	36.00	57.60
Sums		40.25	50.40		36.02	57.85

The market information is identical in the two cases, but the comparison of the two "groups" is dominated in the first case by the *steel* unit and in the second case by the *needles* unit.

The obvious answer to this defect is to introduce weighting so that each price enters the summation with an influence such as it had in the actual market situation. So far, so good. But another complication is encountered immediately. It is necessary to avoid any process that will lead to Σp_0q_0 in the base period and Σp_1q_1 in the given period, since any comparison of these two aggregates gives a measurement of the influence, not of changing prices alone, but of changing prices and quantities together. Clearly, we want a weighting factor that comes from the actual market situation but one that will at the same time give us an aggregate in which the quantity element, so to speak, is held constant while the price element varies. Let us recognize at once that the last statement involves a basic contradiction since price variation in the market is always associated with quantity variation. However, if there is to be any solution, there must apparently be a compromise on varying weights in order to obtain a measurement of price variation by itself. At once, one sort of compromise is logically rejected whereas another deserves at least a certain amount of exploration.

The compromise to be rejected is one that is widely accepted by students of index numbers, namely, to weight prices (both p_1 and p_0) by a set of *normal* quantities (say q_a). Suppose that the price index were to measure cost-of-living changes for a certain group in the community (for example, wage earners and clerical workers in a large city). The q_a might then be a set of quantities hypothetical in char-

acter but of course not too far removed from the amounts actually consumed by these people for the periods in question. The quantities might even be those which this group had consumed in some fairly recent period, being therefore real consumption for them though not q_1 or q_0. The weights could, in fact. have been taken from the actual budget study of the U. S. Bureau of Labor Statistics, 1934–35–36. The reason why this set of weights q_a must be rejected on logical grounds is that the weights q_a are, by hypothesis, neither q_1 nor q_0 and therefore do not reflect the importance of the p_1 or p_0 prices faced by the group in question in periods 1 and 0. By this reasoning, if quantities are to reflect the importance of p_1 or p_0 they must not go beyond the historical record of the actual quantities which this group of persons purchased at the prices of the market. It must not be forgotten that this argument is stated as the exact logical position. It does not deal with approximations, nor with the difficulty that is often met in real life in finding the actual quantities associated with a given set of prices. The position taken here is that a logically exact solution must first be sought and that approximations should be considered later only in those situations where, for adequate reasons, the logically exact solution is unobtainable. Doubtless the above rejection will be challenged by some students of index numbers since it runs contrary to almost universal practice, but the position is still logically unassailable.

The above argument has said that we reject q_a as weights for either p_1 or p_0 when the period a refers neither to 1 nor to 0. Exact weights for expressing the importance of prices must be either q_1 or q_0 because these q's are the only ones which are *causally* related to these p's. Consider then the set of quantities of the base period (q_0) as weights measuring the importance of p_0 and (approximately) of p_1. The q_0's are the actual quantities that the group of consumers purchased in the base period, at base prices p_0. Now calculate $\Sigma p_0 q_0$, and the result is the actual cost of the base-year budget (the historical record). Next price the same base-budget quantities at prices p_1. To be sure, these quantities were not purchased at prices p_1 since p_1 reflects the change (p_0 to p_1) in the prices at which q_0 were purchased. This new value aggregate is $\Sigma p_1 q_0$ or the hypothetical cost of the base-year budget at given-year prices. The concept is not too far afield from our problem. Now the ratio of these two value-aggregates, namely $\Sigma p_1 q_0 / \Sigma p_0 q_0$, does in a sense give a measurement of the influence of price change (period 0 to period 1), for it applies the two sets of prices to the base-year budget. It is therefore a first tentative measurement of the price influence in V_{01}. The above ratio will be referred to hereafter as L_{01}, after Laspeyres, the man who first proposed this form of index. A

second measurement, and one logically similar to L_{01}, can be obtained by using the budget quantities q_1, this time obtaining two sets of aggregates Σp_1q_1 and Σp_0q_1. Then the ratio $\Sigma p_1q_1/\Sigma p_0q_1$ (now called P_{01} after Paasche, its originator) takes a position of equal status with L_{01} as a measurement of the price influence in V_{01}. In general, L_{01} and P_{01} will not have the same numerical value, and the question arises which is to be preferred. The answer is: Neither. One rests on just as solid a logical foundation as the other. What is to be done under the circumstances? Two alternatives present themselves:

1. If L_{01} and P_{01} do not differ too greatly, accept either or some sort of compromise between them (such as an average) as the best obtainable measurement of the price influence in a value ratio V_{01}.
2. If L_{01} and P_{01} differ widely, it may be necessary to reject both as measurements of price change. This alternative is equivalent to admitting that the problem cannot be solved in terms of the procedures developed above. The book will later develop this point in greater detail and clarify the circumstances in which it may be necessary to reject such methods as these on the ground that they are too inaccurate for practical use.

2. Difficulties in Summing Actual Quantities

When the problem is that of measuring quantity variation for the whole group of commodities, the difficulties with unweighted sums are even greater than with prices. For, whereas significance can sometimes be attached to a sum of prices, Σp_0, since the items added are always money, it cannot so easily be attached to a sum of quantities since quantities have no common unit. It is impossible to add tons, yards, flasks, square feet, etc. But, as with prices, a solution may be attempted by weighting, if it is possible to find a measurement of the *importance* (weight) of a given quantity. The truth is, of course, that the importance of q is measured by p and we are dealing with the same problem as before only approaching it from a different angle. Thus we can measure the quantity of goods consumed (in a family budget, for instance) by measuring it in value terms; Σq_0p_0 measures the value of the budget in the base period, and Σq_1p_0 measures the value of the given-year budget at the same base prices p_0. Let $L_{01} = \Sigma q_1p_0/\Sigma q_0p_0$, the symbol L_{01} here referring to quantity variation, being the same symbol as used above to represent price variation. This is then a measurement of the factor of quantity change (the Q_{01}) in V_{01} when the quantity change is determined by valuing both base- and given-period quantities by *base* prices. The analogy with prices may be continued by calculating a quantity ratio P_{01} by valuing both base- and

given-year quantities by *given-year* prices; thus $P_{01} = \Sigma q_1p_1/\Sigma q_0p_1$, and in general P_{01} will not be equal to L_{01}. Again as with prices, there are two alternatives: (*a*) if L_{01} and P_{01} do not differ too widely, accept either or some sort of cross between them as a measurement of the quantity factor in V_{01}; (*b*) if they do differ widely, consider that this kind of solution of the problem has failed and abandon it.

The preceding pages have attempted to ferret out possibilities of measuring the separate but total price and quantity elements of a value ratio by a summation process. And no acceptable solution was found, in terms of direct and simple summation, for either prices or quantities alone. But if weighting is introduced and if the prices (or quantities) in each period are weighted by a logical weighting factor that in effect holds quantities constant while allowing prices to vary (or vice versa), the result is a pair of total value figures the ratio of which gives an answer of sorts to the question asked: What has been the part of the value change V_{01} attributable to price influence alone (or to quantity influence alone)? That the answer is not too completely satisfactory has been emphasized, for in each comparison one element has involved multiplying the price of one period by the quantity of the other period and these two magnitudes are not causally identified in the market. Nothing particularly novel lies in what has been said above, unless it is the emphasis placed on certain logical relationships between the elements used in the solution, that is, the p's and q's. The refusal to accept as weights the p and q values of any other period so long as the solution refers to periods 0 and 1 is a case in point.

4.4. MEASURING THE TOTAL PRICE AND QUANTITY EFFECTS BY MEANS OF PRICE AND QUANTITY RELATIVES

1. Separating the Price and Quantity Effects

Since p_1/p_0 measures price change without any quantity influence and q_1/q_0 measures quantity change without any price influence, and since both are expressed in percentage terms, it follows that some sort of summation or averaging process applied to a set of these relatives should provide a measurement of total price or total quantity effect.

2. Obtaining the Total Price or Total Quantity Influence by Simple Averages of Relatives

Various methods of averaging these relatives have been used to measure the total influence, that is, the new price or quantity plateau of the given period relatively to that of the base period. The question

what average to use is, in the writer's judgment, of much less importance than the question whether it shall be weighted or unweighted. Index number literature, however, has given a great deal of attention to the question whether to use the arithmetic mean, median, or geometric mean in averaging relatives. But much of the basis of choice between them has disappeared with respect to many of our modern index numbers. When an index number is based on anywhere from one hundred to several hundred different commodities, the choice between these three averages is largely a matter of indifference. The arguments in support of each of the three averages can be indicated in a few words, and any detailed discussion will be avoided. Edgeworth [1] was the most famous and persistent champion of the median, and his strongest defense related to the fact that the relatives averaged are a sample of the total number possible and that the median is less affected by irregular and unusual relative values than are other averages. The index numbers he had in mind usually comprehended a small list of commodities, and his argument had pertinence in such a situation, but it would lose much of its cogency when applied to an index number of 800 items, such as the U. S. Bureau of Labor Statistics wholesale price index. The choice between arithmetic and geometric means has hinged on the nature of the frequency distribution of price relatives. In defense of the geometric is the thought that the distribution is skew with minimum values of zero and maximum values unlimited. There is also the thought that variations in price relatives are variations in percentages and as such they follow a geometric law. This idea no doubt influenced Jevons [2] to use a geometric mean for his famous index number in 1863. The arithmetic average is associated in thought with symmetry of distribution form and in particular with normality of distribution. But these arguments over which average to use have had no prominence in the literature since at least the mid-1920's, probably because of general acquiescence in the view that the arithmetic and geometric averages are of about equal stability when applied to index number relatives. The result is, at least in the United States, that the arithmetic average is almost universally employed.

Now as to unweighted averages. Carli's index number [3] in 1764 was constructed as a simple average of the price relatives for three commodities, grain, wine, and oil, and the result was used to measure the total price effect over the 250-year period from 1500 to 1750. In

[1] See, for example, *Papers Relating to Political Economy*, Vol. 1, Papers H and I.

[2] See *Investigations in Currency and Finance.*

[3] See p. 4.

present terminology his formula was $P_{01} = \frac{1}{3} \sum^{u} (p_1/p_0)$ where p_1 is 1750 prices and p_0 those for 1500. Interestingly enough, the history of index numbers shows a number of applications of Carli's procedure. Probably the most important example of it was the wholesale price index number of the U. S. Bureau of Labor Statistics as it was compiled from the early days of the century until it was revised as a result of Mitchell's study for the bureau in 1915.[1]

Nevertheless, a simple arithmetic average of relatives has not held unchallenged sway as the appropriate form of average. In fact, one of the first important index numbers on record was constructed by Jevons in 1863 to measure the fall in the value of gold after the gold discoveries in California and Australia around 1850. And Jevons not only used the simple geometric mean of price relatives, that is, $P_{01} = \sqrt[n]{\overset{n}{\Pi} (p_1/p_0)}$ but also stoutly maintained the necessity in this instance of using the geometric rather than the arithmetic average. Jevons had the average yearly prices of thirty-nine commodities for the years 1845 to 1862 inclusive. His base price p_0 was the simple arithmetic average of the six yearly price figures 1845 to 1850 inclusive. Then the price for each of the eighteen years in his study was taken as a relative to this 1845–50 average, and the geometric mean of these relatives was finally taken. Jevons' procedure, especially with some refinements to measure importance (weights), has been strongly recommended since his time and has been used for published index numbers, notably by the British Board of Trade.[2]

3. Weighted Averages of Relatives

Simple averages of relatives can by their neglect of weighting be very wrong. Consider again an example of steel and darning needles:

Commodity	Unit	Prices (dollars)		Price relatives
		0	1	
Steel	Ton	40.00	50.00	125
Darning needles	Package	.25	.50	200
				Sum 325
				Average 162.5

[1] Bull. 284.

[2] *Ibid.*, pp. 68–69.

The simple average of relatives indicates that there has been a 62.5 per cent increase in the price level of these two commodities, but it is clear that darning needles have influenced this result out of all proportion to their influence in any real market, because the value of all darning needles produced in a year would be at most a few hundred dollars, whereas the value of a year's product of steel would run into millions. Of course, no competent index number maker would make this particular error; and index numbers constructed as unweighted averages of relatives have often introduced weighting indirectly by their selection of commodities. For example, to increase the importance of semifinished steel in the index there could be two price relatives instead of one—one for plates and one for shapes. But it is generally agreed among students of index numbers that this indirect and informal system of weighting is not enough.

The moderns have challenged the neglect of weighting in the Carli and Jevons procedures. The one basic change that the Mitchell study of 1915 introduced into index number practice in Washington was weighting. For the wholesale price index whose defects brought about the Mitchell study was an unweighted average of relatives, and after Mitchell's recommendations were put into effect it became a weighted aggregate that is easily transformed into a weighted average of relatives.

There are plenty of examples of exceptional price relatives in which the particular price relative is not very important. Furthermore, it is generally true in economics that as price increases quantity decreases. In Mitchell's study of prices made for the War Industries Board during World War I he found a single price relative of several thousand per cent. Of Jevons' thirty-nine commodities, the price relatives of which he used to measure the fall in the value of gold, he found three exceptionally large relatives in 1862. These three, for cotton prices, were 349, 263, and 315, whereas the remaining thirty-six relatives ranged in value from a low of 74 to a high of 167 with an average (simple arithmetic, to illustrate) of 116.1. By including the three cotton relatives the average was increased to 130.9. Since Jevons was not using weights he handled this situation by excluding the three exceptional relatives. His average, the geometric, of the thirty-six relatives was 113.4, and had he included the three high ones it would have been 122.7.

Necessity of Value Weights. Granting that price and quantity relatives should be weighted, we are required to find a proper method of weighting. There is the general proposition to start with that the influence of a particular relative on the average should be equivalent to

its influence on that segment of the economy to which the group of relatives belong. To obtain a basis for an answer let us suppose that we are concerned with your cost of living. The whole set of prices and quantities which form the fundamental data from which will ultimately come the answer to the question, "What has been the change in your cost of living?" is the actual set of p_0, p_1, q_0, and q_1 for all the commodities that you consumed in the two periods. Now ask what is the measure of the importance to you of a given p_1/p_0. Let the commodity be *shoes*, and let p_1/p_0 be 1.25, and suppose that you wish to know how important to you is a 25 per cent increase in the price of shoes. Clearly the answer has to do with your total expenditure on shoes, that is, with a value figure (pq). But what total expenditure? Shall it be p_0q_0 or p_1q_1, that is, either of the two amounts that you actually spent in the years 0 or 1, or shall it be some hypothetical expenditure, say p_aq_a, that is somehow defined as normal or typical of your expenditure on shoes.

But p_aq_a Values Must Be Discarded. It seems clear to the writer that a final judgment can be given as to the use of p_aq_a as a weight for p_1/p_0. It was stated in the very beginning[1] that this book seeks the answer to a historical question, namely, "What has been the change in an actual price level (or quantity level) from period 0 to period 1?" The question refers to history, and the exact answer, if it can ever be found, must be derived from the historical record. If this is so, and it must be if our task is as defined above, then the value expenditure p_aq_a for some period other than the base or given period of a stated comparison has nothing to do with our problem *in any direct sense.* It will therefore not be legitimate to introduce p_aq_a into the solution *except as an approximation* to some more exact solution based on the historical facts of the two periods in question. Many published index numbers, of course, use value weights for price relatives which remain fixed over a long period of years; that is, they use p_aq_a. A notable example is the U. S. Bureau of Labor Statistics cost-of-living index. In fact, the federal cost-of-living index has had only two sets of budget figures in its entire history, the first based on a 1919 budget study and the one now in use based on a budget study for years 1934–35–36. These circumstances make the currently published cost-of-living index not an exact measurement of what it purports to be, or *should purport to be,* that is, a measurement of the actual change in the cost of living between two stated periods, say 1945 and 1946, but rather a measurement of the changed cost of a particular budget, namely, that of a given

[1] See p. 17.

group over the years 1934–35–36, on the assumption that this budget has been priced at actual prices of 1945 and 1946. In other words, the index published is based upon a hypothetical budget,[1] a market basket for the years 1934–35–36 and the prices of 1945 and 1946 rather than a measurement based on data of the markets for 1945 and 1946 alone. The purpose here is not to deny all validity to the existing index but to call attention to its approximate and somewhat unreal character. It might still be the best attainable in the existing situation—but more on this subject later.

What *pq* Values to Use. If price relatives are to be weighted by value weights, the values must come from the actual price and quantity records of the two periods compared; this is the unassailable logical position. Let us return to the example of your shoes, where $p_1/p_0 = 1.25$, and remember the historical fact that you spent a certain sum of money, p_0q_0, on shoes in the base period. Now if the price change p_1/p_0 is multiplied by this quantity, that is, $(p_1/p_0)(p_0q_0)$, it will give the sum of money p_1q_0, which is clearly sufficient to purchase in the period 1 the same quantity (q_0) of shoes as were actually purchased in the period 0. Perform this operation for every budget item common to the two years (the N_{01} commodities of earlier notation), and add the results, and we obtain a weighted sum of price relatives $\Sigma(p_1/p_0)p_0q_0$, reducing to Σp_1q_0 or the sum of money necessary to buy the full set of q_0 quantities at the p_1 prices. But the weighted average of price relatives then becomes this figure divided by Σp_0q_0, or the base-year expenditure, and this average is the average price change when p_0q_0 weights are used. Remember now that Σp_0q_0 represents a historical record of fact, the actual expenditure in the base period on N_{01} commodities. But Σp_1q_0 represents a hypothetical sum of money necessary at p_1 prices to purchase that same base-year budget, the q_0's. The ratio of these two sums of money, which is the same magnitude as the weighted average of the price relatives above, is clearly a measurement of the changed cost of the budget q_0's as prices change from p_0 to p_1 and may be accepted as a tentative measurement of the price level change P_{01} in the defined situation. We have then $P_{01} = [\Sigma(p_1/p_0)p_0q_0]/[\Sigma p_0q_0] = [\Sigma p_1q_0]/[\Sigma p_0q_0]$. This is the same Laspeyres formula we reached on p. 20, and it turns out that weighting prices by base quantities gives the same algebraic result as weighting price relatives by base-year values. The result has the virtue of

[1] Hypothesis is, of course, not eliminated by confining the data to the p's and q's of the base and given periods, nor is there any implication here that making hypotheses is wrong. The trouble is that the hypothesis in question has no part in the comparison we wish to make.

providing a measurement of price change through the use of data drawn only from the two periods of comparison. The weakness of the formula lies in the fact that it omits the quantity data of the given period. And the question naturally arises: Why should the q_0's have been used to the exclusion of the q_1's?

Remembering that the last result was obtained by starting with the actual base-year expenditure Σp_0q_0, let us reexamine the problem by starting with the actual given-year expenditure Σp_1q_1. First, note that the actual amount paid for *shoes* in period 1 was p_1q_1 and since prices had changed in the ratio p_1/p_0 from period 0 it is clear that a sum equal to p_0q_1 would have purchased the same shoes q_1 in the period 0. Therefore, we say that p_0q_1 is a measurement of the importance of the price change p_1/p_0, since p_0q_1 dollars will purchase the quantity q_1 at base prices p_0, just as $(p_1/p_0)p_0q_1$ or p_1q_1 purchased the quantity q_1 at p_1 prices. Following this line of reasoning for each of the commodities involved in the binary comparison, and then taking ratios of the two aggregates, we have for the whole set

$$\frac{\Sigma(p_1/p_0)p_0q_1}{\Sigma p_0q_1} = \frac{\Sigma p_1q_1}{\Sigma p_0q_1}$$

as a measurement of the price change between 0 and 1 based upon the weights p_0q_1 for price relatives, or based upon the given-year quantities q_1 as weights for prices. But this result turns out to be the same as the Paasche formula for prices; and thus, as did the Laspeyres formula, shows an algebraic identity between a weighted average of relatives and a weighted aggregative This time the average-of-relatives formula is weighted by the values obtained by multiplying base prices and given-year quantities, whereas the aggregative weights prices of both base and given periods by the given-year quantities q_1.

Thus what seems to be sound reasoning with respect to weighting of price relatives brings two answers, the Laspeyres and Paasche indexes of price change; neither can be considered preferable to the other for logical reasons, and both were obtained wholly in terms of the data of the two periods of comparison. Furthermore, the two answers are the identical two that were reached earlier when approaching a solution by appropriate, logical weighting of prices.

In this choice of value weights for weighing the price relatives p_1/p_0, the point has been stressed that no p and q values may be logically introduced into the value weights except those of base or given period; and the weights p_0q_0 and p_0q_1, designated by Fisher as weights I and II for the price relatives p_1/p_0, meet this criterion. It may be noted

briefly, since the literature of index numbers sometimes emphasizes the fact,[1] that two more value weights for price relatives are obtainable from the p's and q's of base and given periods, but must be discarded for other reasons. They are what Fisher called value weights III and IV for weighting p_1/p_0 or p_1q_0 and p_1q_1. Remember that when p_1/p_0 was multiplied by base-year values p_0q_0 the product was p_1q_0. This result was obtained for each individual commodity, and the two values p_0q_0 and p_1q_0 gave us two costs of a given quantity q_0 at the two prices p_1 and p_0. The figure p_1q_0 was a measurement, so to speak, of the increased expenditure on shoes caused by price change alone. Then a summation of the two sets of values for N_{01} commodities gave two value aggregates—a common budget (q_0) priced at two sets of prices. This calculation has an understandable element of reality as a measurement of change since it gives two costs of a common basket of goods, so to speak, at the two price levels. An analogous situation held for p_0q_1 weights. But it is easily seen that no such result emerges when p_1/p_0 is multiplied by p_1q_0. For now the base quantity q_0 is multiplied first by p_1, which gives an interpretable result, then multiplied by p_1/p_0, that is, $(p_1/p_0)(p_1q_0) = (p_1^2/p_0)(q_0)$, which is a quite meaningless result. So any summation of such results for N commodities is also meaningless. A similar difficulty follows the attempt to weight the same price relatives by p_1q_1. The conclusion follows, therefore, that of Fisher's four systems of weighting involving base- and given-period prices and quantities, namely p_0q_0, p_0q_1, p_1q_0, and p_1q_1, only the first two have any reason for extended consideration. The last two must be discarded at once.

The Weighting of Quantity Relatives q_1/q_0. The line of reasoning employed in finding logical weights for price relatives can be followed through directly in seeking weights for the quantity relatives q_1/q_0. The importance of a quantity change is measured by the money value to which it is to be applied. Therefore weights must be values. Certain money values must be discarded, such as p_aq_a, because values of other periods are irrelevant when the problem is to measure the level of quantity change between periods 0 and 1. Values compounded of the p's and q's of base and given periods are, as applied to quantity relatives, q_0p_0, q_0p_1, q_1p_0, and q_1p_1. These values in this order would be listed in Fisher's terminology as weighting systems I, II, III, and IV for quantity relatives. By application of the same basic steps in reasoning as used above for price relatives, we find that weighting systems I and II for quantity relatives produce results interpretable

[1] See Fisher, *The Making of Index Numbers*, Chapter 3.

in realistic terms whereas weightings III and IV do not. In summary, weightings I and II give weighted averages of quantity relatives which reduce as follows:

Weighting I:

$$\frac{\Sigma(q_1/q_0)q_0p_0}{\Sigma q_0p_0} \qquad \frac{\Sigma q_1p_0}{\Sigma q_0p_0} \quad \text{Laspeyres (for quantities)}$$

Weighting II:

$$\frac{\Sigma(q_1/q_0)q_0p_1}{\Sigma q_0p_1} \qquad \frac{\Sigma q_1p_1}{\Sigma q_0p_1} = \text{Paasche (for quantities)}$$

The reductions to the aggregative form show that two weighted measurements of average quantity change have resulted: one in which both quantities have been weighted by base prices, and one in which both have been weighted by given-year prices. Since neither set of weights is preferable to the other, the two formulas stand on an equal footing; neither is exact, since certain data in a two-way comparison have been omitted, but both are good in that they have not gone beyond the data of these two periods.

CHAPTER 5

Emerging Formulas for Binary Comparisons

5.1. THE TWO BASIC MEASURES, L AND P

The analysis in Chapter 4 has brought forth two measures of price change (L and P) and two of quantity change (L and P) based upon the observational data of an actual historical situation. They will be numbered formulas 1 and 2. Both L and P can be expressed either as the ratio of two value aggregates or as a weighted average of relatives. Each of these results was obtained by an independent attempt to approach the problem of measuring a change in price or quantity levels from a logical standpoint, that is, by asking and answering the question: what is a proper measure of the importance of a price (quantity), or, alternatively, what is a proper measure of the importance of a price (quantity) relative? The two separate approaches through actual prices and quantities and through relatives, led to the same answer, a Laspeyres formula associated with base-year quantities (prices) and a Paasche formula with given-year quantities (prices). These two formulas have equal validity in measuring the percentage of change in price or quantity levels between two periods. They occupy a central position in what follows. But they always differ in any real situation inasmuch as some quantities and prices always change from one period to another and therefore the weights of L do not agree with those of P.

The question is, then, what to do about this difference? If L is not greatly different from P either might be accepted as a satisfactory approximation to the desired measure. The literature of index numbers, however, contains several other formulas that bear close kinship to L or P and that have been advocated by different students of the subject. Specifically, there are crosses of L and P; there are crossed-weight formulas, in which the L and P weights are crossed; and there is the very-much-used and strongly defended fixed-weight aggrega-

tive, in which the weights are presumably not too far removed from the weights of L and P. To these formulas let us devote a brief consideration.

5.2. CROSSES OF L AND P

Naturally one of the first compromises to come to mind, once L and P are recognized as good and as of equal status as measures of price and quantity change, is to take a mean of the two. Two such means have been proposed, one arithmetic and the other geometric They are, in the simplest of L-P terminology, $\frac{1}{2}(L + P)$ and $\sqrt{L \cdot P}$. Or, more specifically, in p-q terminology and in aggregative form, they are, for prices:

(*a*) Arithmetic cross:

$$\frac{1}{2}(L + P) = \frac{1}{2}\left[\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} + \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}\right] \qquad [3]$$

(*b*) Geometric cross:

$$\sqrt{L \cdot P} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \cdot \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \qquad [4]$$

The corresponding formulas for quantities are obtained by interchanging p and q in the above Of these two types, the arithmetic was suggested by Drobisch in 1871 and by Sidgwick in 1883. It was recommended by Bowley in 1901 because it averaged the *inferior* (L) and *superior* (P) limits of the index number. It is one of the two methods most strongly recommended by Walsh.[1] The geometric cross, in more recent years known as the Ideal, a name given it by Fisher in 1920, was suggested by Walsh (1901), Fisher (1911), Pigou (1912), and Allyn Young (1921).[2]

5.3. CROSSED-WEIGHT FORMULAS

Of crossed-weight formulas two have also been advocated, an arithmetic and a geometric cross. Written in p-q notation, they are, for prices:

[1] See Walsh, *The Problem of Estimation*, pp 104–105.

[2] References to the earliest appearances of the Ideal are: Walsh, *The Measurement of General Exchange Value*, p. 429; Fisher, *The Purchasing Power of Money*, p. 418; Pigou, *Wealth and Welfare*, p. 46; and Young, in *Quart. J. Economics*,

(*a*) Arithmetically crossed-weight aggregative:

$$\frac{\Sigma(q_0 + q_1)p_1}{\Sigma(q_0 + q_1)p_0} \qquad [5]$$

(*b*) Geometrically crossed-weight aggregative:

$$\frac{\Sigma\sqrt{q_0 \cdot q_1}p_1}{\Sigma\sqrt{q_0 \cdot q_1}p_0} \qquad [6]$$

Again the corresponding formulas for quantities are obtainable by interchanging p and q. The arithmetically crossed-weight formula is often known as the Marshall-Edgeworth formula because of its advocacy by these two outstanding authorities. It is highly recommended by Fisher and is one of the two best "practical" methods recommended by Walsh.[1] The geometrically crossed-weight aggregative, formula 6, is one of Walsh's choices for the best formulas in theory [2] (of which he names four and of which the Ideal is one).

5.4. THE MOST FREQUENT "PRACTICAL" COMPROMISE, THE FIXED-WEIGHT AGGREGATIVE

The formula that is probably used more frequently than all others combined in the construction of published index numbers of prices and quantities today is the fixed-weight aggregative. One of the earliest index number formulas, it was proposed in 1812 by Arthur Young, in 1822 by Lowe, and in 1833 by Scrope, all of England. In current notation, this formula for prices is

$$\frac{\Sigma p_1 q_a}{\Sigma p_0 q_a} \qquad \frac{\Sigma(p_1/p_0)(p_0 q_a)}{\Sigma p_0 q_a} \qquad [7]$$

where q_a are quantities that in some way measure the *importance* of the prices. None of the early writers were very specific as to what period q_a (or $p_0 q_a$) represented. As Walsh says of Arthur Young's weights, they referred to "the relative total exchange values of the classes, *in general,* at no particular period."[3] The modern vogue of

August 1921. See Fisher, *The Making of Index Numbers,* pp. 240–242, for a historical note.

[1] Walsh, *The Problem of Estimation,* p. 105. Walsh's other best "practical" method is the arithmetic cross of L and P discussed above, formula 3.

[2] *Ibid.,* p. 102.

[3] Walsh, *The Measurement of General Exchange Value,* p 536.

this formula without doubt dates from its advocacy by Mitchell in his classic study of index numbers in 1915.[1] In his conclusions (Bull. 284, p. 113), he says with respect to his general-purpose index series that the best weights to apply to a weighted aggregate of actual prices "are the average physical quantities of the commodities bought and sold over a period of years."

5,5. THE QUALITIES OF THESE FORMULAS—IS THERE A BEST FORMULA?

The status of the L and P formulas, numbers 1 and 2, has been made evident already. On purely logical grounds, and it should be noted that as yet there is offered no other test of the quality of an index number, the two formulas are good, and they are equally good. They are good because each uses data from periods 0 and 1 for the single comparison of these periods and they use no other data. All the observational information that goes into the index is tied together in the causal complex out of which came these two sets of prices and quantities. The two formulas are equally good also because each uses only part of the weight data related to the two sets of prices (quantities), and base weights have no more justification than given-year weights. Or the two may just as well be said to have equal shortcomings since neither uses weight data for both periods. They thus become complimentary measures and lead us to thoughts of compromise.

There is an argument of considerable potency which throws some light on the difference between these two formulas. It is that the Laspeyres formula for prices overestimates the price change whereas the Paasche formula underestimates it. The Laspeyres formula, it will be remembered, measures the change in price levels by pricing the base-period set of quantities, at both sets of prices, p_1 and p_0, and then taking the ratio of these two aggregates. The base aggregate $\Sigma p_0 q_0$ sufficed to purchase the actual set of quantities q_0. Now if the sum of money represented by $\Sigma p_1 q_0$ were provided for the same group of consumers (or producers, or other specified group) it would enable these consumers in period 1 to buy the identical set of commodities in the identical amounts q_0 as in period 0, and therefore, on the hypothesis that they would actually buy this base set of commodities in quantities q_0, the ratio L measures the percentage change in cost *due to change in prices*. That this percentage measure is *too high* is due to the fact that in an assumed free market the consumers, in possession of the

[1] Bull. 284.

money sum $\Sigma p_1 q_0$, would shift their purchases in such ways as to improve their situation. In other words, the percentage change in income measured by L would explain mainly changes in prices but also, in part, changes in quantities. To illustrate, for a stated set of consumers, if the $\Sigma p_0 q_0$ includes an item of \$1.00 that purchased sweet potatoes in amount $q_0 = 2$ bushels at $p_0 = \$.50$, and if $p_1 = \$3.00$ it is certainly true that many consumers, being provided with \$6.00 in year 1 to buy 2 bushels of sweet potatoes, would shift part of this sum to other, more satisfactory ways of spending some of the \$6.00. This reduction in the consumption of any commodity when the price becomes inordinately high is a phenomenon of the market too common to require further emphasis. So $\Sigma p_1 q_0$ enables the consumer to *raise* his standard of consumption, and the Laspeyres formula, therefore, overestimates the price rise to him on the *basis of his base-year budget.*

An analogous argument, starting with the actual expenditure $\Sigma p_1 q_1$ in the given year, shows that the sum of money $\Sigma p_0 q_1$ would have made possible the purchase of the set of quantities q_1 in the base period and that, therefore, on the assumption that this set had actually been purchased in the base period, the Paasche formula measures the price change in this situation. But, again, in a free market with individuals freely choosing, any price changes will be met by appropriate adjustment of quantities and therefore Paasche's formula *underestimates* the influence of price change (in the direction base to given period), since the base aggregate $\Sigma p_0 q_1$ measures the influence not only of price change but also, in part, of quantity change—always, in this instance, with reference to the given-year budget q_1's. This discussion of overestimation and underestimation by L and P has been carried on in terms of price indexes. The argument can without difficulty be extended to quantity indexes.

Note that, although the Laspeyres formula, by this argument, *overestimates* the price change, *it does so with reference to the base quantities* q_0 and the Paasche formula *underestimates* the price changes *with reference to the given-year quantities* q_1. It is still possible that in a given situation the Laspeyres formula can report a *lower* price change than the Paasche formula; that is, the situation may well occur where $L < P$. In fact, it happens repeatedly. For example, in Fisher's *The Making of Index Numbers* [1] he calculates index numbers for his thirty-six commodities for the years 1914 to 1918 on the 1913 base. For the Laspeyres and Paasche formulas he gives the following:

[1] See p. 503.

FISHER'S PRICE INDEXES

(BASE 1913)

Formula	Year				
	1914	1915	1916	1917	1918
Laspeyres	99.93	99.67	114.08	162.07	177.87
Paasche	100.32	100.10	114.35	162.05	177.43

The figures show that the Laspeyres index was below Paasche's in three of the five years.

The question arises now whether a compromise between L and P, or between the elements with respect to which they differ, is legitimate. That is basically a question whether either can be accepted as accurate enough for the purpose in hand. If a particular application of the index requires that its value be known accurately to within two points, either L or P could be accepted with some confidence if $L = 150$ and $P = 152$. But if in the same situation $L = 100$ and $P = 200$, doubt is cast upon the measurement as such because the two equally good estimates differ so widely, and therefore an average of them would also be of doubtful value. This matter of accuracy of measurement will be considered in detail in Chapter 6, and we proceed to judge compromise formulas in the remainder of this chapter on the assumption that a compromise is legitimate.

The status of the L and P crosses and of the crossed-weight formulas follows directly from the considerations upon which the two basic formulas were judged; that is, the new formulas will be judged according to the extent to which they use the complete historical record. If neither L nor P is exact, if the logical reason for this lies in the partial character of their coverage of the observational data, and if each compensates for the deficiency of the other, then some sort of an average or cross of the two is better than either one alone. Any of the formulas 3 to 6 thus partakes of the combined qualities of L and P and is without the shortcomings of either of these indexes taken singly, simply because each of the four formulas utilizes all the data of both periods of comparison. It may be worthwhile at this point to note that from the full data of the two periods of comparison the following four aggregates can be calculated, Σp_0q_0, Σp_0q_1, Σp_1q_0, Σp_1q_1, and that the index of either price or quantity by formulas 1 to 5 can be calculated

directly from these aggregates. The aggregates appear explicitly, of course, in formulas 1 to 4, and formula 5 can easily be transformed as follows (for prices, for example):

$$\frac{\Sigma(q_0 + q_1)p_1}{\Sigma(q_0 + q_1)p_0} = \frac{\Sigma q_0p_1 + \Sigma q_1p_1}{\Sigma q_0p_0 + \Sigma q_1p_0}$$

Formula 6 is not obtainable directly from these four aggregates since it uses the geometric mean between the weights of the two periods, but it does utilize all data of both periods. If any of these four formulas, 3 to 6, is preferable to the others the preference will not be based upon the kind of hypothetical grounds thus far considered but rather on some other test of acceptability, such as ease of calculation or of understanding. On this basis some students shy away from geometric means because it is often stated that the user of index numbers, in general not a technician, will not understand a geometric mean as well as he understands an arithmetic one; and that the calculator can produce the arithmetic mean by ordinary arithmetical processes but must use logarithms for a geometric mean. These arguments, for what they are worth, lead to a choice of formulas 3 and 5 over 4 and 6. It was the calculation argument that persuaded Walsh to select formulas 3 and 5 as "unequivocally the best practical methods." [1] The present writer is not too greatly impressed with either the "understanding" or the "calculation" reasons for choosing index number formulas. If other things are equal, well and good. But if an improvement can come through a more intricate method of calculation, at least the value of the improvement should be compared with the extra cost of calculation before a decision is made. As for the "understanding" issue, we all use automobiles, telephones, radios, and many other modern gadgets without knowing too much about how they are made.

The fixed-weight aggregative, formula 7, faces a wholly different judgment from that given to the previous six when it is tested by the sort of logic applied to the others. Formula 7 uses weights that never, except by accident, belong to either the base or given period of a particular comparison. Indeed, Mitchell said, referring to the wholesale price index, that the weights should be "the average *physical* quantities bought and sold *over a period of years*." [2] The support of this formula has been very great among competent students of index numbers for many years, and that support continues to the present day. The formula has the prestige of this wide acceptance and is used

[1] Walsh, *The Problem of Estimation*, p. 105.

[2] Quoted above, p. 34. Italics added.

in the construction of many important current index numbers in the United States. It is well, therefore, that its claims for status be given careful attention. It is, to start with, an answer to the demand that prices should be weighted, that their influence on the final result should be a measurement of their importance in the situation being studied The argument can be put into other words, and the formula has often been defended in these terms, namely, that the average price change over a period is measured by the changed cost of a fixed bill of goods when priced at the two sets of prices, the fixed bill being such as might have been typical of these two periods. To be sure, the first makers of index numbers who sought quantity weights for price indexes did not find a plethora of material.

The first interest in change through time has as a matter of history been an interest in price change, not quantity change, at least among the index number makers. The result was that collections of prices came earlier than collections of quantities on anything like the same scale. Price data on a large scale began in the United States with the collections of the famous Aldrich committee of the United States Senate in 1893, and a governmental agency was soon established (later becoming the U. S. Bureau of Labor Statistics) to maintain up-to-date collections of prices. The collection of quantity data on a comparable scale probably began as a result of the demands for quantity information on industry-wide bases during World War I, and much of it was done by private industrial agencies. These accumulations of quantity data have continued with accelerating volume ever since, the government entering the picture both in assisting private and business agencies and in becoming itself an original collection agency This growth in the collection of quantity data has doubtless been influenced to a greater extent by an interest in the quantity data for its own sake and in the construction of quantity indexes than by the desire to provide weights for price indexes. Whatever the reason, quantity data nevertheless began to accumulate in considerable amounts during World War I, and this tendency has been accelerated as time has passed, both in the intensity with which given fields of production were covered and in the spread to new fields The current numbers of the *Survey of Current Business*, monthly statistical output by the U. S. Department of Commerce, with their literally thousands of time series, are a mark of the growth of time series data on quantities in the United States.

Now that there really is a vast volume of time series data on quantities as well as on prices, it is legitimate to ask: Why does the use of formula 7 persist? The whole answer is not a simple one, but one very

important aspect of it lies without question in man's hesitancy to break away from established methods. When governmental bureaus, working usually with budgets that will not stretch over all the tasks that might be undertaken, have worked out a set of technical procedures to do a particular job and have obtained a permanent and continuing means of financial support for it, it is not surprising that weighty reasons must be brought forth for changing these methods, especially, as would be true here, when the new methods would involve additional costs. It is often true that basic changes in methods of research must wait until a crisis has developed which shows the inadequacy and weakness of the older methods. Thus it happened that the weaknesses inherent in formula 7 showed themselves prominently only when the exceptional changes developed in both prices and production as a result of total war and when the indexes were challenged on the ground of inaccuracy as to the measurements that they were designed to provide.

The basic logical fault of this formula lies in its weights which, for either prices or quantities, belong to neither of the two periods involved in any binary comparison. For example, the revised wholesale price index of the U. S. Bureau of Labor Statistics which followed Mitchell's study used weights of 1909, the only data then available. There was a later shift to 1919 weights. Then through some dozen years prior to 1937, the weights were revised every two years. In the 1937 revision there was return to fixed weights obtained from data of the years 1929–30–31, and these weights were still employed in 1949.

The Federal Reserve Board's index of industrial production was first constructed in the middle twenties with weights from the year 1923; in the revision of the 1940's new weights were taken from the year 1937, and the latter were still being used in 1949. The cost-of-living index of the U. S. Bureau of Labor Statistics was first constructed in the early 1920's with weights obtained from a workingman's-budget study of 1919. There has been one revision, a second and similar budget study for the years 1934–35–36. The index had only these two sets of weights up to 1949, but in that year a third revision was begun. When this revision is finished a completely new set of weights will be available. The piecemeal revisions that have occurred from time to time in all these indexes are a sign that the inadequacy of fixed weights is recognized. Whenever any of these indexes compare price or quantity levels for two adjacent years of recent history, say in the post-World War II period, and employ weight data of years as far in the past as 1930 or 1935, it is reasonable to assume, upon the basis of the argument of the previous pages, that the results may contain a large error.

All these judgments passed upon our selected list of seven formulas are logical judgments. They have been based on the question whether the formulas keep close to the historical record, and number 7 has been condemned because it goes outside that record. To be sure, there are other tests of quality, or of accuracy, and the next chapter will be devoted to them. Meanwhile, one other comparison of the formulas will be of interest. Fisher in his last book, *The Making of Index Numbers*, classified a great number of formulas into seven classes of quality as: (1) worthless, (2) poor, (3) fair, (4) good, (5) very good, (6) excellent, and (7) superlative. For the moment we need not know the full basis upon which he determines quality, but we might note that by his tests of quality the Ideal formula, our formula 4, rates as of highest quality and therefore will fall in class 7 above. The seven formulas are rated as follows in this classification:

Formula 1, Laspeyres base-weight aggregative—very good.
Formula 2, Paasche given-year weight aggregative—very good.
Formula 3, arithmetic cross of L and P (Drobisch)—superlative.
Formula 4, geometric cross of L and P (Ideal)—superlative.
Formula 5, Marshall-Edgeworth arithmetic crossed-weight aggregative—superlative.
Formula 6, Walsh geometric crossed-weight aggregative—superlative.
Formula 7, fixed-weight aggregative—fair.

It is not surprising to find the central four formulas, which depend upon the whole data of both the periods of a comparison, classed as superlative; to find also that the two basic formulas, L and P, out of which the others are compounded, are considered very good; and, finally, to find that formula 7 is rated below all the others in the list But now we must seek a further and objective means for testing accuracy and not depend upon the authority of Fisher's or any other person's name. Some of the possibilities will be explored in the next chapter.

CHAPTER 6

How Accurately Can P_{0k} and Q_{0k} Be Measured?

6.1. THE DATA FOR COMPARING TWO PERIODS i, $i = 0, k$

If any two periods i are being compared there exist three value aggregates $\sum^{N_i} p_i q_i$, $\sum^{N_{0k}} p_i q_i$, and $\sum^{n_{0k}} p_i q_i$, $i = 0, k$ for each one. The first summation, over N_i, covers *every* commodity existing in period i in the defined field of study. Summation over N_{0k} involves just those commodities that exist in both periods 0 and k, that is, binary commodities. Obviously N_{0k} is a different number for every different pair of years. Finally, n_{0k} is a selection from N_{0k} done on a sampling basis. Recall that the first value aggregate represents the sum total of all *values* in the defined field for any period i but that the second and third aggregates cannot be defined until two periods have been paired for comparison, since the summation over N_{0k} involves all commodities found in *both* periods. And it is to be expected that N_{01} for periods 0 and 1 will be a different number from N_{02} for periods 0 and 2 since, in general, the binary commodities for periods 0 and 1 will differ in some degree from the binary commodities of any other pair of periods. Finally, the summation over n_{0k} is a part of the summation over N_{0k} and, if n_{0k} is selected on terms demanded by sampling theory, the results give a representative sample. This means, for example, that in studying any property or quality of the N commodities, such as price change, the n commodities can be used to measure the same property but the result will be subject to errors of sampling.

Each of the value ratios $V_{0k}(T)$, $V_{0k}(N)$, and $V_{0k}(n)$ [1] that can be calculated from these value aggregates is then a measure of value change between period 0 and period k. The differences between the three that were pointed out in Chapter 3 are important for the index number problem, and a clear understanding of them is vital to a knowl-

[1] See Chapter 3.

edge of just what index numbers measure and what they do not measure. First, $V_{0k}(T)$ is the ratio of *all* values in the given period to all values in the base period. It is the ratio of two nonhomogeneous value aggregates—nonhomogeneous in the sense that the numerator contains values for some commodities not represented in the denominator and the denominator contains values for commodities not represented in the numerator. The commodities that are found in only one of any pair of periods are the *unique commodities* of Chapter 3.

Obviously no price or quantity relatives can be calculated for unique commodities, nor can two sets of value weights such as discussed earlier be calculated for them. $V_{0k}(T)$ will hereafter be referred to as the value ratio for full but nonhomogeneous data. All the value data of the two periods are present in the ratio, but the nonhomogeneous elements will not be usable in the methods of measurement based upon formulas 1 to 7. $V_{0k}(N)$, on the other hand, contains *all* the data and *only* the data of the two periods for which price and quantity relatives can be calculated and for which two sets of value weights can be obtained; that is, it is based upon the data of all binary commodities. $V_{0k}(N)$ may therefore be called the value ratio for complete homogeneous data. In this ratio every commodity present in either numerator or denominator has both a p_0 and a p_k and also a q_0 and a q_k. Finally, $V_{0k}(n)$ can be spoken of as a value ratio based upon partial but homogeneous data. It is based upon a partial list of the binary commodities, and the selection is presumably made in accordance with the requirements of good sampling.

At the risk of undue repetition, the above contrasts are stated once more in different terms. In each case a valuation is made of baskets (B) of commodities. With the ratio $V_{0k}(T)$, there are *two* separate baskets and they contain partly different sets of commodities: B_k contains N_k commodities, and B_0 contains N_0 commodities. Both these baskets may have the same kinds of meats and cereals, but one can contain buggies and not automobiles, the other automobiles and not buggies. The ratio $V_{0k}(N)$, on the other hand, does not require two baskets; it needs only one, B_N. The commodities it contains will be present in quantities q_k at prices p_k in period k and in quantities q_0 at prices p_0 in period 0. The particular goods in this basket could be obtained, of course, by taking the content of either of the first pair of baskets B_0 or B_k after having removed from it all those commodities that are not found also in the other basket. Finally, the ratio $V_{0k}(n)$ has only *one* basket, B_n, and its set of commodities can be obtained by dumping some commodities from basket B_N. To be sure, the dumping process should be based on rational sampling considerations.

In summary, $V_{0k}(T)$ contains all *existing* data of the two periods but is a nonhomogeneous value ratio. $V_{0k}(N)$ is homogeneous because it contains only binary commodities; they all are present in both periods. A binary comparison is therefore possible. Furthermore, for binary commodities it is complete. It contains *all* such commodities. Its homogeneity is obtained by neglecting the unique commodities of each period. Finally, $V_{0k}(n)$ is homogeneous but partial. It *samples* the complete list of binary commodities in B_N.

6.2. THE PRICE AND QUANTITY LEVELS WHICH P_{0k} AND Q_{0k} ARE DESIGNED TO MEASURE

In Chapter 3 the index number problem was defined formally by the relationship

$$V_{0k} = P_{0k} \cdot Q_{0k}$$

or as the task of separating a value ratio into the two components, price effect and quantity effect. The same chapter called attention to the three value ratios that were described in some detail in Section 6.1. The question now is: With which value ratio is index making concerned? The answer to this question as it has been worked out in practice has always dealt with $V_{0k}(N)$ or its sampling equivalent, $V_{0k}(n)$, but it may well be that the problem of which an index number is a solution centers upon $V_{0k}(T)$. The present writer believes that the latter proposition is the true one and that the great issues of practical import which have lain behind the growth of the classical index number literature of the past are issues that concern the ratio $V_{0k}(T)$.

Let us examine briefly some of the facts of the record. Carli [1] was concerned with how imports of silver from the Americas had affected price levels over a period of 250 years. We are interested here not so much in his method of measurement as in the historical situation that brought about his study of the problem. Carli observed that prices in monetary terms were higher in his day (1750) than they were reported to have been at the earlier date (1500). One can easily imagine that his interest in the problem reflected a widespread opinion that prices had risen. It would be a quite misguided stretch of the imagination to assume that Carli, or the economists of his day, thought that only those prices had risen which he included in his index (three: grain, wine, and oil). More probably, these were the only three for which prices were obtainable for the two dates and they must have been considered as

[1] See Chapter 2, p. 6. See also Mitchell, Bull. 284, p. 7; Walsh, *The Measurement of General Exchange Value*, p. 534.

representative of a general situation. In short, Carli's thinking dealt with $V_{0k}(T)$ even though his measurement dealt with $V_{0k}(n)$.

Consider another historic example, Jevons' article of 1863, entitled "A Serious Fall in the Value of Gold Ascertained." He actually measured the increase in price of about forty commodities, and upon this increase he *ascertained the fall in the value of gold.* But when he speaks in these terms, he is thinking of gold in *all* its monetary uses, not just as applied to forty commodities. In other words, Jevons, like Carli, was comparing two historic situations and he was concerned to find out whether gold as money had declined in value. *All money* had declined in value, and Jevons was tracing some of its social consequences. This again is a problem of $V_{0k}(T)$, that is, of the total situation.

Fisher's equation of exchange furnishes another illustration of this important concept. Fisher writes $MV + M'V' = PT$, where $MV + M'V' = \Sigma pq$, the latter term meaning exactly what it has meant in this book, namely, the total value of all transactions in some defined field, that is, $\sum^{N_i} p_i q_i$. Then $\Sigma pq = P \cdot T$, where "*P* then represents in one magnitude the *level of prices* and *T* represents in one magnitude the *volume of trade.*" [1] One needs only to take the ratio of two such *P* values of Fisher to have an index number of *price change*, that is, a P_{0k} as the term has been used here. And, again, the ratio of two of Fisher's *T* values gives the corresponding index number of quantity change. But notice that both P_{0k} and Q_{0k} here arise conceptually out of two value summations that in the terminology of this book can be none other than $\sum^{N_0} p_0 q_0$ and $\sum^{N_k} p_k q_k$; hence Fisher's price and quantity *levels* lead to index numbers (price and quantity change) that come from $V_{0k}(T)$. The correct statement of Fisher's problem is

$$V_{0k}(T) = P_{0k}(T) \cdot Q_{0k}(T)$$

It can be demonstrated with equal ease that current uses of index numbers are intended to deal with a similar real situation. There is always the thought of a price level as it actually existed in period 0 and another price level as it existed in period k. Their ratio is the price component for a value ratio $V_{0k}(T)$. Current cost-of-living studies support this thesis. When wages are tied to cost of living the wage earners are comparing their current situation with a past one, and when costs of living rise they prepare to demand step-by-step increases in their wages, to return their status presumably to that of the earlier date.

[1] Fisher, *The Purchasing Power of Money*, pp. 27 and 20–21. Italics added.

Always they compare $\sum^{N_0} p_0q_0$ with $\sum^{N_k} p_kq_k$, and the cost of living in which they are interested is without question the one that measures the price component of $V_{0k}(T)$.

A business periodical of current date reports that industrial production is still moving downward and that further declines are expected. The index is quoted as dropping four points. There can be no doubt about what the businessman is thinking in this situation, assuming that the index accurately reflects his views. It is that the scale of total business activity in the field covered by the index is declining, since $Q_{0k}(T)$ is declining, it being one of the two components of $V_{0k}(T)$.

If these examples are typical, as it is believed they are, then any price or quantity index that is presented as a solution to one of these real-life problems must, if it performs its task with exactness, measure the price or quantity element in a total-value change. Hence the definition of the index number problem as given in Chapter 3[1] may now be stated in a more definite manner and with greater precision as

$$V_{0k}(T) = P_{0k}(T) \cdot Q_{0k}(T)$$

The significance of this statement becomes clearer when the seven formulas of previous pages are recalled and when it is remembered that each of them, being an average of relatives, is therefore a result obtained only by operations on the more restricted body of data contained in $V_{0k}(n)$. Any attempt to measure the accuracy of modern index numbers must therefore, in some manner, take account of the difference between $V_{0k}(T)$ and $V_{0k}(n)$ or between the price and quantity components of these two ratios. This is because the problem arises from $V_{0k}(T)$ but all the modern solutions are associated with $V_{0k}(n)$. To this matter the next section will be devoted.

6.3. THREE COMPONENTS OF ERROR IN $P_{0k}(n)$ AND $Q_{0k}(n)$

Index numbers P_{0k} and Q_{0k} based on modern methods are therefore constructed from the commodity list of $V_{0k}(n)$. The data cover a sample of binary commodities. If they are designed to solve a problem of measuring the price or quantity factors of $V_{0k}(T)$, they are therefore subject to an error of measurement, and this error has three components:

1. A formula error.
2. A sampling error.
3. A homogeneity error.

[1] See p. 16.

The formula error arises from the fact that there is no universally accepted formula that will measure the price change or the quantity change of a given body of data with exactitude. The formulas L and P, for example, are equally right and, thereby, also equally wrong. There is no choice between them. Even the crossed formulas and crossed-weight formulas do not all report the same measurement of price or quantity change. All these formulas cannot be right, and a criterion for selection among them is obviously needed.

It is this particular component of error that has received the greatest attention in the literature of index numbers. Fisher stated in 1923 that the problem of formula error had for all practical purposes been completely solved It was possible, in his judgment, to measure the price change for a given body of data with almost complete exactness. As he said, the error in the Ideal formula "seldom reaches one part in 800."[1] Fisher's opinion indeed did not meet with 100 per cent agreement. English students of index numbers disagreed violently in several reviews of his book, and there was plenty of disagreement among the American statisticians as well.

The sampling error of any price or quantity index based upon any of the formulas 1 to 7[2] arises from the fact that the formula operates upon the commodities list of $V_{0k}(n)$ rather than upon that of $V_{0k}(N)$. Assuming formula accuracy for the moment, then a $P_{0k}(N)$, for example, would give an exact measurement of change in *price levels* for the *complete list of binary commodities*. If, on the other hand, the index number maker works with the more limited data of n binary commodities rather than with the complete list, it is to be expected that his result, $P_{0k}(n)$, will differ in some degree from $P_{0k}(N)$. This difference is a sampling error, and there are well-established methods of estimating it.

The third component of error is a homogeneity error. If it were possible to obtain an exact measurement of it, this magnitude would represent the difference, again assuming complete formula accuracy, between $P_{0k}(T)$ and $P_{0k}(N)$. Or, it is the difference between solving the real problem which is put up to the statistician in the first instance (that is, breaking up $V_{0k}(T)$ into its price and quantity components)

[1] *The Making of Index Numbers*, p 228. The validity of his method of estimating this error can be questioned, but he undoubtedly had in mind what is here called formula error, that is, the ability of the formula to report the true value of what has here been called $P_{0k}(n)$.

[2] The comments to follow will apply without modification to almost any formula that has been or will be proposed for actual use. No restriction is therefore implied in limiting the discussion of sampling errors to formulas 1 to 7.

and solving the problem which the statistician has always tackled (that is, that part of the complete problem which covers only the homogeneous data of the binary commodities of the two periods). No measure of the homogeneity error is available in the literature of index numbers, and indeed this error probably never has been directly formulated in the terms here given. There is no doubt, however, that the basic problem of the difference between full data and complete homogeneous data has been in the minds of students of index numbers. Fisher's article "The Total Value Criterion: A New Principle in Index Number Construction" is a clear reflection of the distinction between these two value ratios, and his *total value criterion* is precisely $V_{0k}(T)$.

6.4. HOW TO MEASURE THESE ERRORS

1. Formula Error

Fisher's Reversibility Tests for Consistency. The formula error, as indicated above, arises from the fact that different formulas report different measurements of price and quantity change even though they hold equal rank from any logical point of view. It is to be noted that formula error has no relevance to the distinction between n and N (this being the area of sampling error) or to that between N_{01} and either N_0 or N_1 (this being the area of homogeneity error). Formula error deals wholly with the properties of a given body of data, the data, of course, from which the index number is to be constructed. This in our terminology is the data of $V_{01}(n)$, that is, the sample body of price and quantity data out of which we construct a value ratio for partial but homogeneous data. It may be said that formula error deals with internal aspects of the sample data of the comparison.

Fisher has probably contributed more than any other writer to the analysis and measurement of formula error. His book *The Making of Index Numbers* was devoted almost wholly to a study of this matter. He measured formula error by the two tests, time and factor reversal. He reasoned that any formula to be accurate must maintain time consistency. It should be independent of the base. If P_{01} equals 125 then P_{10} should equal 80. This test was not one of Fisher's creations. It had been known and its use advocated for many years. If the test is satisfied

$$P_{01} \cdot P_{10} = 1$$

and if it is not satisfied there is a joint error. Fisher called the error joint since there was no basis for assigning its parts to the forward (P_{01}) and backward (P_{10}) applications of the formula. The percentage measure of this joint error is then

$$E_1 = P_{01} \cdot P_{10} - 1$$

The factor reversal test, which originated with Fisher, arises from the argument that a formula which is right as to prices should be equally right as to quantities. Put in other terms, the formula should correctly factor the value ratio. Therefore, the test says,

$$P_{01} \cdot Q_{01} = V_{01}$$

and of course in Fisher's use of the test he was concerned with factoring the value ratio of the given body of data upon which index numbers are to be calculated. Stated in our terminology, this means that

$$P_{01}(n) \cdot Q_{01}(n) = V_{01}(n)$$

or the test is to be applied to the sample data of binary commodities. The failure of a given formula to meet this test produces another joint error, according to Fisher—joint in the sense that it cannot be distributed logically to the price and quantity factors. In order to express this joint error as a positive or negative percentage error, to maintain similarity with E_1, he defined it as

$$E_2 = \frac{P_{01} \cdot Q_{01}}{V_{01}} - 1$$

Fisher used these tests, of course, as a basis for choosing among formulas, and the formulas that he classified as superlative [1] were those which either exactly or very nearly satisfied those tests. They had little or no joint error. Indeed, Fisher's selection of the Ideal formula as logically best in all instances was due largely though not wholly to the fact that it has no joint error by either of these tests. He classified as superlative those formulas that most closely agreed with the Ideal. The critics of Fisher's book were probably more upset by his position on formula error than by anything else. Many of them had already endorsed the Ideal formula as best, notably Pigou and Bowley, but when Fisher held that this formula had measured price change to an accuracy of 1 point in 800, that is, ⅛ of 1 per cent, many disagreed. As said above, Fisher's support of the Ideal as best was only in part due to its ability to meet both these tests exactly. He further advocated the formula on the ground that it measured price or quantity change by utilizing *only* the data of the two periods compared and by utilizing *all of them*. It is the judgment of the writer that the time and factor reversal tests are good supplementary methods to help in

[1] See p. 40

selecting formulas, provided that certain prior logical conditions are met. Of these conditions the one that the formula shall use only the p's and q's of the two periods compared is absolutely basic. A test of formula error that is related to this condition will be recommended shortly. Meanwhile, a brief comment upon another and independent proposal for insuring formula accuracy, the famous Konus condition.[1]

Konus' Condition. Konus published an article in a Russian journal in 1924 dealing with the measurement of change in the cost of living, but it was not until January 1939 that the appearance of an English translation of the article made Konus' views generally available to English and American students. Konus' problem was to measure the true change in the cost of living, and his approach differed in one fundamental respect from the traditional one and from the one employed in this book. Whereas the traditional approach defines a *field* of study, in this case the cost of living of some specified group, for example, wage earners in a certain community, the Konus procedure was to select *two* groups of such persons, one at each of the two periods or two places, the criterion of selection being that these two groups had enjoyed the same standard of living. If two such groups could be found, then the measure of the true cost-of-living change between the two periods is the ratio of their two expenditures, in our terminology, V_{0k}. Konus maintained that if a certain condition were satisfied this equality of standards of living would be sufficiently exact. This condition, stated for periods 0 and 1, now famous as the Konus condition, is that

$$\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_1}$$

It has been demonstrated[2] that this condition does not guarantee that the value ratio V_{01} will lie between the two true indexes $I_{01}{}^0$ and $I_{01}{}'$ based, respectively, on the base- and given-period budgets but will always lie outside them when the two standards differ. The trivial case of identical standards does not ordinarily occur in real life and could not be identified if it did. Since the Konus condition does not guarantee identity of standards in the two situations it does not make possible an exact measurement of the true cost-of-living change; we must seek elsewhere for further tests of accuracy of formulas applying to sample data of the traditional two periods, and, in general, it will be certain that the two budgets differ. In other words, V_{01} always con-

[1] See Konus, *Econometrica*, Vol. 7, pp. 1–29; Mudgett, *Econometrica*, Vol. 13, pp. 171–181.

[2] See Mudgett, *Econometrica*, Vol. 13, pp 171–181.

tains an element of quantity change Q_{01} as well as a P_{01}. Note that Konus sought to identify a situation in which Q_{01} equals unity and P_{01} therefore equals V_{01}.

The D Test for Formula Consistency. D is here defined as

$$D = L - P$$

where L and P are the initials of the Laspeyres and Paasche formulas. It is less a measure of formula error than it is a measure of the lack of agreement between two formulas that have equal status on logical grounds as measures of price or quantity change. It is, therefore, a measure of formula consistency in reporting a given price or quantity change and can be used under the following conditions: If for some purpose it is sufficient to measure price change accurately to within one or two percentage points and if D is then less than two, either L or P would be satisfactory; but if in this same situation D had a value of, say, twenty points, then it would seem that we do not have any satisfactory measure. The correct answer in this circumstance may be that we have no way of obtaining a more accurate measurement and it may be necessary to abandon the project of measurement altogether. The argument upon which this judgment is based may be presented as follows· If we wish to measure price (or quantity) change between two years and if we have available sample data of prices and quantities for each period, then we can measure average price change (1) by pricing the base quantities at both sets of prices and summing the results to $\sum^{n_{01}} p_1q_0$ and $\sum^{n_{01}} p_0q_0$ and taking the ratio of these two values, getting L_{01}; or (2) by carrying out the same operation on the q_1's, obtaining P_{01}. Either of these two results is better than a ratio of two value aggregates calculated for some set of quantities q_a that belongs to neither base nor given year, because the quantities q_0 and q_1 are the ones causally associated in the market with p_0 and p_1, and q_a has no such causal connection.

Therefore, either L or P can be accepted as a standard for measuring the accuracy of any formula that goes beyond the p-q data of the two periods for any of its price or quantity materials (in this case weights). But L and P stand on equal grounds: they are equally good or perhaps equally bad. Suppose that in some hypothetical example $L = 80$ and $P = 82$ If a difference of two points is of no consequence to any application of the index number, then either, or some cross, can be accepted as the best available measure of this price change $D = 2$, and the inconsistency of the two formulas is inconsequential. Suppose, however, that $L = 80$ and $P = 180$. If a choice has to be made here as to which of these shall be used to measure a change in

cost of living in order to form a basis for new wage levels, it is easy to see that the conflict in interest between employer and workman cannot be settled in a way satisfactory to both. One will accept P, the other L, but neither may be satisfied with any kind of compromise between them. In other words, with D equal to 100, the index number maker may well say that the problem must be given up rather than solved with a mean of two such widely divergent results. The uncertainty of the measurement is so great in this example as to render undesirable any attempt to put it in quantitative terms.

The above comments are equivalent to saying that at times we can accept any of the formulas 1 to 6 as satisfactory index number measurements but at other times the differences between them make any attempt at measurement foolhardy. It is easy to see wherein the difference D arises. L and P differ only in their weights, one employing base-year weights, the other given-year weights. Clearly, if D is small it must mean that the q_0's and the q_1's are not too divergent. How shall we be assured of this? One good guess is to compare periods (and places) that are close together in time (and space). It looks as though we might compare this year and last year with tolerable accuracy. The farther apart in time the two periods under comparison, the greater the inconsistency and the less the accuracy. Furthermore, the mere measurement procedure may introduce a spurious accuracy, an actual numerical value, but if L and P are 100 points apart there is no way of telling which has more accurately measured the change in price level. But the fuller development of this idea must be delayed for a later paragraph. Meanwhile let us look to methods of measuring sampling error.

2. Sampling Error

The sampling error of index numbers arises from the fact that calculations are based on a set of n commodities found in the two periods of the comparison and this set is used to represent the whole list of N common commodities. For the purpose of considering this error factor, the price or quantity change based on N represents complete accuracy, say $I_{01}{}^N$. This magnitude is then without sampling error. The index based on n, however, is an estimate $(I_{01}{}^n)$, and $I_{01}{}^n - I_{01}{}^N$ is therefore the sampling error. In the usual statistical sense it is a variable, and, therefore, for all possible samples of n that could be taken from N there is a frequency distribution of these errors. We need only a knowledge of some of the properties of this distribution in order to gain needed insight into the accuracy of any determination of I_{01}. This knowledge is readily available from modern statistical theory. Since almost all

index number formulas in practical use are averages of relatives, and since the entire seven discussed in this book are of this variety, we are here concerned with the sampling distribution of an average, and the error of sampling is none other than the standard deviation of the sampling distribution of an average. Several features require explanation in a full statement of the sampling error of an index number formula (average). They are:

1. Error of an average of a random sample.
2. Effect of weighting.
3. Effect of sampling from a finite population.
4. Effect of sample stratification

If a random sample of n magnitudes X is taken, the sampling error of $\bar{x}_1$, where $\bar{x}_1 = (1/n) \sum^{n} X$, is

$$\sigma(\bar{x}_1) = \sigma/\sqrt{n}$$

where σ equals the standard deviation of (the population of) the X's. An estimate of σ would generally have to be obtained from the sample data. For index numbers constructed as simple averages of randomly selected price or quantity relatives, it is therefore possible to obtain from the above formula estimates of their sampling errors. It may be noted that the variance (squared standard deviation) of an average varies inversely with the size of the sample, so that, other things being equal, the larger the sample the smaller the sampling error.

Most index number formulas, however, are weighted averages, and it is now almost universally recognized that weights cannot be neglected. Their apparent neglect at times is often compensated by certain devices of *hidden* weighting. If a weighted average is defined as $\bar{x}_2 = (1/\Sigma w)\Sigma wX$, where w is the weight of a given X, and if σ is the common standard deviation of all the X's in the population, then the sampling error (squared) [1] or variance of $\bar{x}_2$ is

$$\sigma^2(\bar{x}_2) = \sum\left(\frac{w}{\Sigma w}\right)^2 \cdot 2$$

Either of these two formulas may be used in its appropriate place, once a value has been obtained for σ. In general, this standard deviation must be estimated from sample data.

A complexity now arises from the fact that samples of index number data are usually not simple random samples. Samples are called *simple random* when for each selection of an item for the sample every item

[1] The squaring obviates the need for square-root signs.

in the population has an equal chance of being selected. The principle of selection is clear; the actual means of getting a simple random sample is often exceedingly difficult. In general, the best that can be expected is a realistic approximation to theoretical requirements. Actual samples of index number data are likely to vary in two respects from the simple random type. First, they are drawn from a finite population, wherefore they may be described as samples without replacement, and, second, they are stratified. Because the samples are without replacement, the selection of one magnitude for the sample, that is, one price ratio, will change the probability of the selection of that price ratio, and also of other price ratios, as later items in the sample. If, for example, the first price ratio chosen among foods at retail is for bread, then, since this item will not be selected *again* the probability of its further selection is zero, whereas probabilities for other foods are at the same time increased.

The principle can be further illustrated by another example. Suppose that the N existing price ratios (the whole population of binary commodities) are ticketed and the tickets placed in a bowl. Then sample selection may be carried out by drawings from the bowl. If the whole sample is to be strictly *simple random* (drawing with replacement), then after each drawing the ticket is replaced in the bowl; but, if the drawing is without replacement, each ticket is held out until the entire sample is withdrawn Obviously if $n = N$ and if the drawing is without replacement, all tickets are drawn from the bowl in each sample and the sampling error reduces to zero since the average of N price ratios is the true average of the entire population (of binary commodities). The effect, upon the variance of the average, of sampling without replacement is indicated by the formula

$$\sigma^2(\bar{x}_3) = \frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right)$$

Here N is finite; if N were infinite, the failure to replace the items as drawn would not affect the probabilities of future drawings. Index number samples are generally samples of this second sort, and so the extra factor $(1 - [n-1]/[N-1])$ may be important in any estimate of the sampling error of an index number. In some of our most important index numbers of national scope this feature of the sampling error has been reduced to very small proportions by taking very large samples. Not only is n large absolutely but also it is large relatively to N. Thus the U. S. Bureau of Labor Statistics wholesale price index has for some years contained over 800 price items. With this number

of commodities it is easy to see that the factor $(1 - [n - 1]/[N - 1])$ may be a very small quantity and may result in reducing the sampling error of the index to a very small magnitude. It might even be possible to say that such a comprehensive index is practically devoid of sampling error.

The last point that needs to be considered in measuring sampling errors of index numbers is that they are practically always stratified samples. It can be illustrated by the cost-of-living index. A wage earner's family expenditure can be classified into several important categories, such as (1) foods, (2) clothing, (3) rent, (4) utilities (fuel, light, etc.), (5) household expenditures, and (6) miscellaneous. A random sample of n commodities could be without any representatives from one or more of these classes, but a stratified sample would have to give proportional representation to each class. This procedure of stratification, if carried out effectively, insures that the sample will be more nearly representative of the entire N commodities just because it requires that every important subclass of commodities will be represented in the sample. The stratified sample is sometimes referred to as a stratified random sample since each stratum is represented proportionately in the sample and at the same time the sampling within each stratum is done on a *simple random* sampling basis. Stratified sampling increases accuracy over random sampling for a given size of sample; that is, the standard error of a stratified sample of size n is less than the standard error of a simple random sample of the same size. In our terminology the variance of a stratified sample may be written

$$\sigma^2(\bar{x}_4) = \frac{\sigma^2}{n} - \frac{1}{n}\sigma_{m_s}{}^2$$

where $\bar{x}_4$ is the mean of the stratified sample, m_s is the true mean of stratum s, and σ_{m_s} is the standard deviation of the means of all strata. Since σ^2/n is the variance of a random sample of n, the formula shows that stratification reduces the sampling variance by the nth part of the variance of the strata. Stratification can be effective, therefore, in reducing the sampling variability of sample averages.

3. Homogeneity Error

The third component that was listed in Section 6 3 is homogeneity error. As was there noted, it arises from the fact that index numbers are calculated from data of binary commodities, whereas the problems that have led to the demand for measurement have arisen from the

totality of commodities in the two periods compared. This distinction is formalized in the proposition that the true price index based upon existing methods of calculation is $P_{0k}(N)$ or the true measure of price change among *all* binary commodities, whereas $P_{0k}(T)$ is the true price index that answers the question how much inflation has taken place. This latter index must be constructed, if at all, from the data of all commodities of both periods, both binary and unique. Then the formal measure of homogeneity error is $P_{0k}(N) - P_{0k}(T)$. As stated earlier, there is no existing method of measuring the difference since no one has proposed a means of measuring $P_{0k}(T)$. Indeed, there has been no previous formulation of the idea of a homogeneity error, unless it has escaped the writer's attention, although history contains many instances of economists' or statisticians' regarding the problem as one of the relationship between two price or quantity levels. Remember Fisher's conception of the meaning of P and T in his equation of exchange.[1] Luckily, however, we are not in the position of being able to do nothing about homogeneity error even though we cannot measure it. We can get some idea of its magnitude in many actual situations, and we can employ methods of minimizing that magnitude.

The *R* Test for Homogeneity. If R is defined for any pair of comparison periods as

$$R = \frac{\text{number of unique commodities}}{\Sigma(\text{unique} + \text{binary})\text{ commodities}}$$

then in any actual situation there always exists a real value for R. For, if the complete data for both periods were gathered, the determination of both numerator and denominator of this ratio would then be only a matter of counting and comparing. And in any actual situation, one could make a pretty good guess as to the make-up of this ratio. Its value, as is quite clear, depends upon the list of unique commodities—that list of commodities found in period 0 and not in period 1 plus those found in 1 and not in 0, all those commodities for which price and quantity relatives cannot be calculated. In symbols the number of unique commodities is $(N_1 - N_{01}) + (N_0 - N_{01})$, and the total number is $(N_1 + N_0)$. Therefore,

$$R = \frac{N_1 + N_0 - 2N_{01}}{N_1 + N_0}$$

R measures homogeneity in this sense:

[1] See p. 11 for quotation from *The Purchasing Power of Money.*

(a) if $R = 0$, there is complete homogeneity.
(b) if $R = 1$, there is complete heterogeneity.

For, if $R = 0$, there are no unique commodities, and $N_1 + N_0 = 2N_{01}$; the numerator equals zero. If $R = 1$, there are no binary commodities; $N_{01} = 0$, whence $R = (N_1 + N_0)/(N_1 + N_0)$. The value of R must always lie between these limits; that is, $0 \gtreqless R \gtreqless 1$.

It must be clearly recognized that R is to be used only as a very general guidepost. As a measure of error, it does little more than indicate how much of the total picture is omitted in current methods of index number construction. But this information is important, for if we know that R is close to zero our current methods are very close to a solution of the real problem of changing price levels. Indeed, if $R = 0$ the problem is solved by current methods; that is, it is solved in the sense that $P_{0k}(T) = P_{0k}(N)$.

What, then, are some of the important causes of the variation in R? Clearly, they center on those influences of an economy that bring about change. If, for example, we define a field of investigation as the people in a compact, easily ascertainable group, all the wage earners in a particular factory for instance, then we can compare price levels for factory hands in this firm for any periods over the life of the firm Surely, if we compare their purchases in 1948 and 1949, there will not be many unique commodities, and R will be close to zero. But if the firm has been in operation for a hundred years and if we try to compare the price levels of today's workers and those of a hundred years ago, it is more than likely that the list of unique commodities will be large. R may then be close to unity, and, no matter how sure we are of the similarity of the N_{0k} commodities which are priced for both periods and how easy it may therefore be to calculate a $P_{0k}(N)$, we must still recognize that this may be a very inadequate measure of a $P_{0k}(T)$, if indeed it has any meaning at all for two periods a century apart. Thus it would seem that the farther apart in time are two periods compared the greater the difficulty of being sure what our measurement represents.

Separation in space creates a similar difficulty. The comparison of price levels for the workers in New York and Newark may be accurate and may be realistic, but extend the range over international borders and difficulties multiply. How can you compare the change in price levels between New York and Warsaw?

6.5. HOW TO CONTROL THESE THREE ERRORS

There have been hints in earlier pages as to what can be done to minimize these errors, but a fuller statement is in order. The difference D between the Laspeyres and the Paasche formulas is a measure of the inconsistency between two equally valid formulas and is explained wholly in terms of changes in weights. For a price index, D will equal zero only when $q_1 = q_0$ for all commodities involved, which of course does not happen in real life. It is clear, however, that the less the change in the conditions of demand and supply between the two periods, the smaller will be D. This fact points to the desirability of making comparisons only of items closely related in space and time, if it is desired to keep formula error to the minimum. Thus we are led to the rule that comparisons of price and quantity change should be made between consecutive periods or areas. This advice, however, will meet with much opposition from many students of index numbers, for it discourages direct comparison between distant periods, which they want and insist must be made. The parity price program in agriculture, for example, made it a requirement in law that certain of farmers' prices should be kept on a parity with those of the pre-World War I period. The law, of course, takes for granted the possibility of measuring changes in price levels between 1910 and years later. In the next chapter a means will be considered for constructing index number series which it is hoped will preserve all that is valid in these distant comparisons. At this point, the only question at issue is how to keep D small, and the sure answer is to make all comparisons between consecutive periods. It is recognized at once that this advice runs into some difficulty when the principle is applied to place comparisons, since the place equivalent of a time series comparison is hardly realistic. For the moment attention is directed only to the fact that formula error may be very low in making any comparison between 1949 and 1948 but in a comparison of 1949 and 1849 may be so great as to destroy all confidence in the measurement. Just as a legitimate sidelight on this issue, it may be remarked that there is very little economic meaning anyway in a comparison of price levels of two periods a century apart. The meaning is even more obscure when we compare price levels in New York and in the Gobi desert.

The means of controlling sampling error is readily at hand after considering the error formulas given above. Each formula contains root n in the denominator, and thus, other things being equal, increase in sampling accuracy follows, step by step, increase in the size of the

sample. The formal rule is that error varies inversely with the square root of sample size. The factor related to sampling from a finite population further indicates an increase in sampling accuracy as n approaches N; that is, this element of accuracy varies with the percentage that the sample size bears to the total population. It is assumed in these comments that the theoretical requirements of sample selection are met. If a random sample is indicated, then the conditions of randomness must be satisfied. Most index number samples are stratified; hence it is assumed that the strata are known and that the sample is actually drawn proportionately from each stratum.

Homogeneity error is controlled by keeping the list of unique commodities small. The method of doing so, as we have seen, is by comparing consecutive periods. An interesting fact is that one procedure controls both the homogeneity error and the formula error at the same time.

As the distance increases between compared periods, there is greater divergence in the means of production and the ways of consumption Changes in resources (the discovery of new or the exhaustion of old), changes in technical methods of production, the growth of inventions, changes in population and in tastes—all these bring about an increase in the number of unique commodities and therefore an error in distant comparisons. In other disciplines allied to index number making, particularly those involving the data of index numbers, there have been attempts to make comparisons over time by reducing divergent items to an ultimate common denominator, as by reducing different foods to caloric contents or by expressing children as percentages of adults These schemes have limitations, as is evident when foods are compared only in caloric content. Modern biochemistry has found many important ingredients in foods besides those which produce heat or energy. All these facts reveal the difficulties inherent in distant comparisons between prices and quantities. And, of course, these difficulties are largely avoided by making the comparisons between consecutive periods. This procedure is the way of controlling both formula error and homogeneity error. It gives assurance that index number theory is capable of producing highly satisfactory results in practice if applied where accuracy is possible and if the sources of great inaccuracy are avoided. It may still be possible to construct a valid procedure for series comparisons and possibly also one for comparisons of distant areas. This problem will be the subject of Chapter 7. Meanwhile we have not quite completed the subject of errors. There are other types of errors, although they will not be given extended attention here.

6.6. ERRORS IN THE BASIC DATA OF INDEX NUMBERS

The kinds of errors which we have not yet considered and which we will give merely passing attention here are those associated with the basic data of index numbers. They involve the accuracy of the original p's and q's. In the presentation of problem and data in Chapter 3, the data were defined as including the complete historical record. The total record for period 0, for example, was expressed as a value summation, $\sum^{N_0} p_0 q_0$. In the entire analysis of the problem to this point it has been assumed that the data of p's and q's were definitive, that is, that there is no error in the record of prices and quantities taken from the market. Now, of course, there is error in this record, particularly inasmuch as prices for any period, say a year, must of necessity be average and error may exist in the individual price quotations and therefore also in the average price for the period. Similarly, errors are to be expected in the quotations for quantities from whatever source they may be drawn. In addition to the quotations of p and q, there is another source of possible serious error in the data gathered. In the discussion of sampling error, certain formulas were given to measure the influence of various sampling features, on the assumption that samples were of a specified variety. The standard sample of data for index numbers, for example, is a stratified sample from a finite population. The formulas measure errors only if the sample meets the theoretical requirements of that kind of sample. But, should a sample fail in this respect, it contains another component of error. Here, then, are two additional sources of error in index number construction: (1) errors in the "observed" p's and q's, and (2) bad sampling methods. It must be emphasized that they are both of very great importance; that their proper handling requires special training; and that this book deliberately omits an analysis of them, since its basic purpose is to stress the other named sources of error, and since the practitioner may often be less adequately acquainted with formula, sampling, and homogeneity errors than with errors in the original data or in sampling methods.

CHAPTER 7

Long-Distance and Series Comparisons

7.1. THE BINARY COMPARISON APPLIED OVER WIDE RANGES OF TIME OR SPACE

It is doubtless true that the most important and most frequent uses of index numbers in modern times are those which are concerned with the interpretation of immediate past history and its relation to the present. Inflation, for example, becomes a problem when it causes someone to find his economic responsibilities more difficult to meet today than yesterday. It may be the wage earner who is faced with higher living costs but no change in his wages, or the businessman who is meeting higher costs of production without a corresponding increase in his selling prices. The previous six chapters of this book have dealt mainly with the measurement features of this problem: how best to measure the change in the level of prices or of production between two situations—the binary comparison. Most emphasis has been placed upon nearby comparisons; we have noted the greater accuracy of comparisons of adjacent periods and the tendency for accuracy to decline with increase in distance or time. Attention must now be focused on one particular application of index numbers that is of much more questionable value. The comparisons involved are those of widely separated situations and, in particular, series comparisons, that is, index numbers for each and all of a series of time intervals over some considerable period of time. We have, for example, both annual and monthly index numbers of wholesale prices in the United States extending from the mid-eighteenth century to the present time. Two questions at once arise concerning series comparisons, namely: Are such long-term comparisons needed, and are they realistic?

1. Are Long-Term Comparisons Desired?

The answer to the first question is definitely an affirmative. Long-distance comparisons have been demanded in the past and they continue to be demanded. Recall Carli's problem of measuring the price increase over a two-and-one-half century period. Two more recent examples are the farm parity price problem and the problem of wages of Ford Motor Company employees in different parts of the world. The former example concerned a law passed by the U. S. Congress in the early 1930's which provided for governmental support of certain farm prices whenever they fell below a position of parity with prices that farmers paid. The law stated that farmers' prices were to be maintained at the same relative position which they had held with respect to these other prices in the years 1910–14. The law, of course, could not be put into operation until the index numbers were constructed for prices received by farmers and for prices paid by farmers, and thus a long-term comparison for the years 1930 and thereafter was involved.

The Ford wage problem arose when the Ford company proposed in the late twenties to introduce a wage policy to provide its employees with the same standard of living wherever they lived, whether in Detroit, London, Berlin, or Warsaw; and there were, of course, Ford agencies in most of the larger cities of Europe. Thus, the vital problem of index number comparisons over space arose, and it here involved such matters as the differences in the cost of living in widely separated areas such as Detroit and Warsaw. The separation was not only geographical but, even more important, it was one of basic differences in the habits of living: different political systems, different resources, different production techniques, different social and economic stratifications. And who will deny that all these variables tremendously complicate the problem of comparisons of prices over geographic space?

The demand for series comparisons arises because modern students of economics, indeed of all the social sciences, find that many of their basic studies of social causation require knowledge of a variable over time, since the variability in which they are interested occurs over time. These comparisons require time series of data—population and production, for example, by years or months, even in some cases by weeks or by days. Immediately the phenomenon of production levels or of price levels arises, and these variables are also wanted by years or months to compare and to correlate with the others. Almost any modern research into the problems of the business cycle involves the need for time series data, and questions of group production or group

price levels bring out at once the need for price and production indexes expressed in series over time. Thus most index numbers of wide import today, such as those dealing with industrial production and with wholesale prices for the whole economy of the United States, are published (1) to supply immediate comparisons with the recent past, and also (2) to furnish the latest figures in the series of continuous data that extend into the more distant past. Indeed, one annual series of wholesale price indexes in the United States goes back to colonial days.

2. Are Long-Term Comparisons Realistic?

To the question whether these comparisons of distant situations are realistic, whether the measurements that we ultimately obtain have any counterpart in the world we live in, the answer is not so easy. In a properly qualified sense it is both yes and no, for there are some realistic features in the comparison of two widely separated periods but it is doubtful whether the limitations of such comparisons are always fully recognized.

Consider the question of changes in wholesale price levels in the United States since Civil War days. All available indexes point to a long-time trend toward lower price levels from the Civil War until about 1896. And very few students of the economic history of the United States over this period would question the reality of this decline. We remember too well that it nearly changed the political complexion of the country. People will also remember the swing to higher price levels during World War I, culminating in a maximum for the wholesale price index in May 1920. Again, there is no doubt about price inflation in the post-World War II period. These directions of movement of price levels have had too much impact upon our daily activities to be doubtful, at least for those who have lived through these times. There are, however, certain features of this historical record about which there can be doubt. For example, the question arises: How do the top levels of prices for the several wars of the period compare? The top level for World War I with that for the Civil War? Or for World War II with that for World War I? Often we hear the statement that prices seek a stable level after each war but that each succeeding level is somewhat higher than the previous one. Many people believe this, and any specific index number published over the years between the two periods of comparison gives an unequivocal answer. But are we sure of this answer? There is no doubt that we possess a numerical figure showing the price level for 1950 relatively, say, to 1896, and some people can still recall examples of wide dis-

crepancies in prices actually paid, for example, eggs at 7¢ per dozen in 1896 and ten times that in 1950. One wonders whether this memory of some detail of the past is made clearer by neglecting many other commodities, say, automobiles or good roads, which may have actually declined in price. One may also recognize that there have been many additions over the years to the commodities we enjoy and many improvements. How shall we, for example, compare costs of transportation between these two dates? In 1896, buggies, wagons, carts, railways, automobiles; in 1950, railways, automobiles, trucks, airplanes. By the later date, the whole picture had shifted. Railways performed a different service in 1950 from that performed in 1896, and this change is in part due to other methods of transportation that have been developed. We are here dealing with what the economist calls the changing state of the arts, one of the most important factors that complicate these comparisons over time.

3. The Difficulties of Measurement

For the moment this disagreement about what we can know about price levels of the past will be resolved by saying that we can know a great deal if only the recent past is involved but that our picture of the past loses sharpness of detail and of outline as the date recedes in time. A nearby comparison is the one that gives us our notions of inflation; and inflation is real to us just because we can compare what we paid today with what we paid yesterday. It is the same phenomenon that was referred to a few lines above in comparing the direction of price movement, which was downward for the years immediately following the Civil War, downward from 1929 to 1933, and upward through the post-war years of the 1940's. But, when we make comparisons that do not involve this process of *having lived through the years in question,* doubts of the reality of index number measurements may arise. Take again the example of comparing wartime price levels. An existing index number of wholesale prices indicates that the maximum level of wartime prices was approximately the same for (1) the War of 1812, (2) the Civil War, and (3) World War I. The question is whether existing methods of measurement of price levels give us any confidence in such comparisons. It is this writer's belief that they do not. We can, with the aid of index numbers, know much about the direction of change in price levels. This direction of change is a phenomenon of movement through time, of comparison of adjacent periods. It is not a matter of the direct comparison of distant periods like 1896 and 1950. When we jump a wide space of years and try to

make direct comparisons without these intervening periods we are on much less solid ground. Maybe we can go so far as to establish order relationship between such price levels, relationships of "more" or "less," without being able to go further and put the "more" or "less" in quantitative terms.

Consider just one more example of the difficulties involved. Suppose that it were desired to compare the standards of living of farmers in America today and in colonial times. An index of the quantities consumed by these farmers, if it could be constructed and if it had any meaning, would report the change in levels of living from colonial times to the present, and this change would be reported as a percentage *increase*. There is little doubt in the minds of most people that farmers enjoy higher living standards now than then, and the index measurement would doubtless support the general impression. But consider the differences between the two ways of living: in the earlier period primitive tools, much manpower, much dependence on uncultivated nature for foodstuffs, clothing, and shelter; today, on the other hand, better food animals through cultivation, selected or developed seeds, for example, through hybridization, the replacement of much direct manpower with greater and greater use of mechanical power, electricity, etc. Then how shall we compare fresh with pasteurized milk, the bread of olden times with the modern variety with vitamins added?

We may leave the issue with the reminder that what the American sometimes calls a higher standard of living is often scornfully repudiated by the French peasant, who is sure that his simpler, *more primitive* life represents a much higher standard of joy and comfort.

These, then, are the kinds of problems with which the index number maker is faced when his measurements are designed for uses other than period-to-period comparisons, and especially when the periods of comparison are far apart in either time or space. But his difficulties are not limited to lack of realism in these comparisons. They are matched, step by step, by uncertainties in the measurements themselves. For we must now recall the nature of the errors in a binary comparison and, in particular, the *formula* and *homogeneity* errors. Recall that increasing values of D, the difference between measurements by the Laspeyres and Paasche formulas, and of R, the ratio of unique to all commodities in a two-way comparison, are practically sure to result from an increasing span of time or space between the two situations compared, because of the greater opportunity of variation in the p's and q's and because of changes in the list of commodities. And when D and R increase there is greater uncertainty about the true value of $P_{0k}(n)$ (related to D) and about the true value of $P_{0k}(T)$ (related to R). This is the back-

ground of information which we possess as we contemplate the prospect of making long-range comparisons of price and quantity levels or of constructing a series of index numbers over a period of time.

7.2. THE SERIES COMPARISON

The long-range comparisons that index number users wish to make are, then, of two kinds: first, the straight two-way long-range comparison, in which the index number users want to compare, for instance, today with 1910 in regard to the parity price problem; and, second, the series comparison, in which they require the whole series of index numbers for every month from 1910–14 to the present time. Whichever kind of comparison it may be, the answer to the question how best to make it is the same, and the discussion in the remainder of the chapter will be framed in terms of the problem of the series comparison.

The question to be answered, then, is how best to construct an index number series. Two preliminary and minor matters may be dealt with first, namely: (1) how important is the selection of a base for the series, and (2) should the index be independent of the base, or should the base be shiftable without error? Following the consideration of these points, the major problem will be to choose between two methods of calculating an index number series, the fixed-base method and the chain method.

1. The Selection of a Base

If one checks through the chapters on index numbers in almost any of the textbooks on elementary statistics, of which a great number have been published in the last quarter century, one will find one prominent feature to be a discussion of the importance of selecting a base for the series. Emphasis is laid upon the selection of a normal period, presumably related to the purpose of the index number. This stress upon choosing a normal period indicates a pattern of thinking that is quite common among textbook writers and, indeed, it must be very common among students of index numbers in general. It seems to be based upon one of those unquestioned bits of lore which have come to be a part of the literature, no more to be doubted than the observed price data themselves; and one wonders where the idea started that the base period must be normal. As a speculation, here are two possible sources. The modern development of index number theory roughly parallels the development of modern business cycle theory, and the business cycle students, as soon as they dealt with empirical observations, had, of necessity, to select a *normal* from which to measure movements towards prosperity or depression. And, of course, many empirical series of

cycle data have been index numbers in the sense in which most people use the term. But there is another source of this modern insistence on the selection of a normal period as base: authority. Probably the most important single piece of literature on index numbers, if we judge by influence upon the later practice of index number construction, has been Mitchell's study *The Making and Using of Index Numbers*,[1] and Mitchell laid great emphasis upon the importance of selecting a proper base. Mitchell's authority has been great among students of index numbers, and very deservedly so, for his bulletin is unquestionably one of the landmarks of all time on the subject. It will not in any way reduce the importance of his enduring contributions to the subject to call attention to one point in his bulletin which will not stand the test of time. If the index number problem is to find out how much the price (quantity) level has changed from one period to another, then the *base* of the index number appears merely as one of the two periods of this binary comparison. Furthermore, if the formula used meets the test of time consistency, then, if P_{01} is known, P_{10} must of necessity equal $1/P_{01}$, and each period has in turn been used as base.

The above result does not follow from any determination that one of the two periods is normal. Nothing more is involved than the comparison of the price levels of the two periods. If P_{01} is calculated the period 0 is the base, and if P_{10} is calculated the period 1 is the base. Which index shall be calculated is a matter of the interest of the index number maker, and his decision has no necessary relationship to the normality of either period.

The base period can always be considered a norm in a limited sense since it provides a level with which the second period is compared. That is, P_{0k} measures the extent to which the price situation has changed from period 0 to period k. If Congress decides, as it did in passing the price-parity law, that certain farmers' prices shall be maintained on a parity with those prices in 1910–14, these years become the base period by legal action and they are a norm for the purpose of carrying out the provisions of the law. The normality here, however, applies to a use of the index number and not to any of its inherent properties as such.

2. Should the Base be Shiftable without Error?

This question is closely associated with the problem of base selection. For, if a particular index number formula is independent of its base, the base can be shifted directly without error. Suppose, for example,

[1] Bull. 284.

that a series of indexes $P_{01}, P_{02}, \cdots, P_{0k}$ is available, and suppose that for some reason P_{45} is desired. If the index is independent of its base, P_{45} may be obtained readily and without error by the simple expedient of dividing P_{05} by P_{04}. If this independence does not hold, then the only means of obtaining P_{45} on the same level of accuracy as the original series is by a direct calculation, returning to the original p-q data.

Some students of the subject demand that an index number shall be independent of its base and thus shall be shiftable directly without error. This insistence is equivalent to setting up a new test of accuracy of index number formulas, the circular test. Formally, the test can be written

$$P_{01} \cdot P_{12} \cdot P_{23} \cdots P_{(k-1)k} \cdot P_{k0} = 1$$

and in this form indicates the source of its name. For the above says that if index numbers are calculated from one period to another over several periods so that the final factor gives the index for the original starting point with the last period, k, as base, then this around-the-circle trip should bring the index back to its starting place. A slightly different form of the test is

$$P_{01} \cdot P_{12} \cdot P_{23} \cdots P_{(k-1)k} = P_{0k}$$

This equation states that the product of the several *links* from 0 to k should equal the directly calculated P_{0k}. If the time reversal test is met, this statement becomes equivalent to the first one. It is easy to see that this test is an extension of the time reversal test.

Before passing any critical judgment upon the circular test, it is well to list the formulas that satisfy it. It is, and can be, satisfied only by unweighted or constant-weighted aggregatives or geometric *averages* of relatives.[1] Thus among the seven formulas that have received consideration in this book, only one, the fixed-weight aggregative, meets the new test. And the fact that this formula is the only one of the seven not seriously recommended from a theoretical point of view gives a rather pronounced adverse judgment upon the circular test.

It is interesting to note the position taken by some authorities on the circular test. Fisher, for example,[2] says that it is theoretically wrong and then devotes twenty pages to proving that it lacks little of fulfillment by practically all his *superlative* formulas. Walsh [3] employs it as

[1] See Fisher, *The Making of Index Numbers*, pp. 274–275, also Appendix 1, Note A, to Chapter XIII, § 4.

[2] *Ibid.*, p. 271; see also rest of Chapter XIII.

[3] *The Problem of Estimation*, see Index; see also his *The Measurement of General Exchange Value*.

a preliminary selector, discarding formulas that do not meet the test but using further bases of choice among formulas that do meet it.

The position taken by this author, in agreement with Fisher, is that the test is theoretically wrong. The objection to it can be explained by reference to either form of the test as given above, in association with the facts that have been developed in the previous six chapters about accuracy of measurement. Consider the first formula. The various steps P_{01}, P_{12}, etc., in going from period 0 to period k conform to experience for that is the direction of movement through time, and measurement of these steps can be highly accurate; but the step P_{k0} represents a backward step that cannot be taken We cannot undo the history of inventions, the discovery of new resources and new commodities, and particularly we cannot undo the exhaustion of certain of our resources; therefore we cannot recreate the conditions of demand and supply, of consumption and production of a long-past period (0) We did not shift directly from the economic circumstances of 1896 to those of 1950, and it is therefore useless to create a measurement that carries the implication that we did so. In short, neither P_{0k} nor its reverse, P_{k0}, provides a measurement that has any precise relationship to actual economic experiences in these two widely separated periods The adjacent links P_{01}, P_{12}, etc., can each be taken with all the accuracy of the best known (superlative) formulas. But their product cannot equal P_{0k} calculated by the same superlative formula since this formula would be subject to all the errors of a distant comparison to which the previous pages have been devoted—errors of formula consistency measured by D and errors of homogeneity measured by R. Thus, if the reasoning of these pages is accepted, the circular test is basically wrong and its satisfaction by any formula is one (at least minor) indication that the formula itself is wrong. This comment should not be taken as a proposal to use the circular test in this negative way to eliminate formulas from consideration. The more basic methods of elimination to which the earlier parts of this book have been devoted are the appropriate ones for this task.

We reach one curious result at this point. Most students of index numbers today appear to accept the fixed-weight aggregative formula as being, for practical purposes, the best, at least if the weights are revised every decade or two; but they also seem with about equal unanimity to insist upon the importance of the selection of some *normal* period as base. But fixed-weight aggregatives are shiftable without error and are therefore no more tied to one particular base than to any other base. In other words, the same value of P_{45} is obtained

by dividing P_{05} by P_{04} and by a direct calculation of P_{45}, when the fixed-weight aggregative formula is used. But remember that constant weights are never right. The correct set of weights for periods 0 and 1 is of necessity different from those for the periods 1 and 2. Satisfaction of the circular test, therefore, cannot be right if constant weights cannot be accepted.[1]

[1] The statement above is correct if all calculation details in such indexes conform to the requirement of constant weights. It is quite possible that calculation procedures are employed whereby subindexes are calculated with fixed weights and then a weighted average of these subindexes is taken to give the final index; and in this whole process the *constancy* of weights can be violated. In this event the statement above about shiftability without error is not precisely correct, although in most practical cases the difference is likely to be very small. For example, suppose that an index is to be constructed as a fixed-weight aggregative

$$P_{01} = \frac{\sum^{n} p_1 q_a}{\sum^{n} p_0 q_a} = \frac{\sum^{n} \frac{p_1}{p_0} p_0 q_a}{\sum^{n} p_0 q_a}$$

the latter term giving the weighted average-of-relatives equivalent of the aggregative. Now suppose that two subindexes are first calculated by the average-of-relatives formula and are then combined into P_{01} The two subindexes will be

$$P_r = \frac{\sum^{r} \frac{p_1}{p_0} p_0 q_a}{\sum^{r} p_0 q_a}; \qquad P_s = \frac{\sum^{s} \frac{p_1}{p_0} p_0 q_a}{\sum^{s} p_0 q_a}$$

Then P_{01} will satisfy the first formula above and the requirement of constant weights, if it is calculated from P_r and P_s as follows:

$$P_{01} = P_r\left(\frac{\sum^{r} p_0 q_a}{\sum^{r} p_0 q_a + \sum^{s} p_0 q_a}\right) + P_s\left(\frac{\sum^{s} p_0 q_a}{\sum^{r} p_0 q_a + \sum^{s} p_0 q_a}\right)$$

$$= \frac{\sum^{r} \frac{p_1}{p_0} p_0 q_a}{\sum^{n} p_0 q_a} + \frac{\sum^{s} \frac{p_1}{p_0} p_0 q_a}{\sum^{n} p_0 q_a}$$

$$= \frac{\sum^{n} \frac{p_1}{p_0} p_0 q_a}{\sum^{n} p_0 q_a} = \frac{\sum^{n} p_1 q_a}{\sum^{n} p_0 q_a}$$

and when the weighting procedure is followed precisely in combining subindexes, the final indexes as well as subindexes can be shifted without error. Additions and substitutions in the commodity list of this index may also affect the shiftability of the base.

3. Fixed-Base vs. Chain Methods of Calculating a Series

Definitions. In theory there are two methods of calculating a series of index numbers, fixed-base and chain. In practice, at least in the United States, the fixed-base is used almost universally. The difference between the two procedures is easily shown in the usual symbolic notation. If we designate the base period as 0 and the successive periods which follow as 1, 2, $\cdots$, k, the full series of fixed-base price indexes will be

$$P_{01},\ P_{02},\ P_{03},\ \cdots,\ P_{0k}$$

where each term expresses the percentage change of the price level for that year upon the 0 base and each is calculated by a direct application of the particular formula to base and given year. The symbolic representation of the chain series is somewhat more complicated, inasmuch as it is necessary first to have every individual link in the chain, such as P_{01}, P_{12}, P_{23}, $\cdots$, $P_{(k-1)k}$. Then the several terms of the series are obtained by multiplication as follows:

$$\begin{aligned} P_{01} &= \text{first link} \\ P_{02} &= P_{01}\cdot P_{12} \\ P_{03} &= P_{01}\cdot P_{12}\cdot P_{23} = P_{02}\cdot P_{23} \\ &\text{etc.} \end{aligned}$$

Comparison of Fixed-Base and Chain by Two Formulas. Each method, as is clear, gives a series of indexes starting with the base year equal to 100. But corresponding later terms in the two series need not report the same measurement and, in fact, can show wide differences, as can be fairly easily judged by writing in p-q notation the results for one or two formulas on both fixed-base and chain methods. Consider first formula 7, the fixed-weight aggregative:

Series Symbol	Fixed-Base Index	Chain Index
P_{01}	$\dfrac{\Sigma p_1 q_a}{\Sigma p_0 q_a}$	$\dfrac{\Sigma p_1 q_a}{\Sigma p_0 q_a}$
P_{02}	$\dfrac{\Sigma p_2 q_a}{\Sigma p_0 q_a}$	$\dfrac{\cancel{\Sigma p_1 q_a}}{\Sigma p_0 q_a}\cdot\dfrac{\Sigma p_2 q_a}{\cancel{\Sigma p_1 q_a}} = \dfrac{\Sigma p_2 q_a}{\Sigma p_0 q_a}$
P_{03}	$\dfrac{\Sigma p_3 q_a}{\Sigma p_0 q_a}$	$\dfrac{\cancel{\Sigma p_2 q_a}}{\Sigma p_0 q_a}\cdot\dfrac{\Sigma p_3 q_a}{\cancel{\Sigma p_2 q_a}} = \dfrac{\Sigma p_3 q_a}{\Sigma p_0 q_a}$

Here we meet the surprising result that fixed-base and chain methods agree. This result is due, of course, to the fact that the formula satis-

fies the circular test. These considerations may explain much of the support for this formula among the experts. The supporters accept fixed weights, not only because the formula then gives the ratio of two costs *for the same bill of goods*, at p_0 and p_1 prices, and is thereby readily interpreted as a measure of price influence alone (since quantities are constant), but also because the identity of fixed-base and chain series gives to the final results a simplicity and an understandability that are satisfying. These properties would be acceptable were it not that the formula is basically wrong in the use of fixed weights instead of the changing weights that arise from the actual price (or quantity) changes in the market. Indeed, the simplicity and understandability are deceptive influences. Simplicity has subtracted something from reality—that is, the quantities q_a have been substituted for the q_0's and q_1's that are causally related to the prices p_0 and p_1—and the understandability detracts from the actual complexity of the market.

Let us now make a similar comparison of fixed-base and chain series with another, better formula, Laspeyres' formula (1):

Series Symbol	Fixed-Base Index	Chain Index
P_{01}	$\dfrac{\Sigma p_1q_0}{\Sigma p_0q_0}$	$\dfrac{\Sigma p_1q_0}{\Sigma p_0q_0}$
P_{02}	$\dfrac{\Sigma p_2q_0}{\Sigma p_0q_0}$	$\dfrac{\Sigma p_1q_0}{\Sigma p_0q_0} \cdot \dfrac{\Sigma p_2q_1}{\Sigma p_1q_1}$
P_{03}	$\dfrac{\Sigma p_3q_0}{\Sigma p_0q_0}$	$\dfrac{\Sigma p_1q_0}{\Sigma p_0q_0} \cdot \dfrac{\Sigma p_2q_1}{\Sigma p_1q_1} \cdot \dfrac{\Sigma p_3q_2}{\Sigma p_2q_2}$

If attention is now centered for the moment on the P_{03} it will be apparent why these fixed-base and chain indexes diverge. P_{03}, when calculated directly as a single term in the fixed-base series, utilizes only the two sets of prices p_3 and p_0 and the base set of weights q_0, despite the fact that the actual price movements between the periods 0 and 3 involved four sets of prices p_0, p_1, p_2, and p_3 and also the corresponding four sets of quantities q_0, q_1, q_2, and q_3. It should be easy to understand this P_{03} because it does have simplicity, but that quality has been obtained by the neglect of most of the economic facts that produced the transition of prices from the 0 to the 3 period.

The chain index P_{03}, on the other hand, looks complicated and it is complicated. It is more difficult to explain to the uninitiated than is the first P_{03} but nevertheless, we must note, it has made use of all the p's of the market and all except the last year's set of q's. This fact

cannot help but make it a more accurate measurement. And while the point is under consideration, it may be well to remark that understandability of a tool of science by the layman is not a consideration to be very strongly urged[1] *unless all other things are equal.* And in this vital case they are not equal. One is reminded, by analogy, of the trite remark that few of us *understand* everything about the mechanics of an automobile but most of us drive just the same. Incidentally, automobiles continue to be improved and the customer continues to know very little about their technical construction. We need the same improvement in index numbers.

The Pros and Cons of Fixed-Base Series. The favorable arguments for fixed-base series are few and are easily stated. They are, in fact, the understandability discussed in the previous paragraphs plus the matter of easy calculation. Nothing further need be said about understandability except that it has its limits and that they are quickly reached. If fixed-base and chain series agree (the case for formula 7), then of course there is no difference in ease of calculation because there are no differences at all. This fact alone may add another *prestige* value to this formula and provide another reason for its extensive vogue. But, in all the remaining six formulas singled out in Chapter 5 as very good or superlative, there are differences between the fixed-base and chain systems and these differences may become very great. The claim that for these formulas the fixed-base system is easier to calculate than the chain system is undoubtedly true. Consider the calculation of the index P_{04} by the two methods, having in each case all the previous terms in the series. If Laspeyres formula is used the fixed-base index is

$$P_{04} = \frac{\Sigma p_4 q_0}{\Sigma p_0 q_0}$$

and the numerator represents the only new calculation that needs to be made. If, on the other hand, the chain system is used then

$$P_{04} = P_{03} \cdot P_{34}$$

and P_{34} requires calculation anew. Here

$$P_{34} = \frac{\Sigma p_4 q_3}{\Sigma p_3 q_3}$$

and two value aggregates are to be calculated in this instance. Also the *new* quantities q_3 must be obtained. There is no doubt that the

[1] The author is fully aware that he is challenging eminent authorities here. All too much of the literature of the subject has been devoted to making index numbers simple and understandable to the public.

chain index takes more calculation than the fixed-base index for this formula, as do, of course, the remaining five formulas of Chapter 5. If there were high correlation between ease of calculation and basic accuracy this favorable feature of the fixed-base series would be important. Unfortunately there is no such correlation and the benefit of ease of calculation can be obtained only by sacrificing more basic considerations.

What are the arguments against the fixed-base system? They are, in fact, those defects that are found in any binary comparison as the distance increases between the periods compared. P_{01} by the standards of earlier chapters will in general be highly accurate, P_{02} less so, and then, as time passes and the base period recedes farther and farther into the past, accuracy will continue to decline. D, the measure of the different reportings of Laspeyres' and Paasche's formulas, can become very large as q_0 and q_k diverge farther and farther, and R, the proportion of unique commodities, can move toward unity. And remember that when a large proportion of the *existing* commodities of two periods is classed as unique, the remaining binary commodities are a correspondingly small proportion of the total, and, therefore, index number formulas, which are applied to the latter, lose much of their ability to report measurements that apply to the whole economic situation, that is, to $P_{0k}(T)$ as defined in earlier pages. No other defect of the fixed-base series need be sought in order to condemn the system utterly, since this particular defect is of such magnitude. Consider what it means in a practical situation. Organized farm groups first supported a principle of price parity (parity with the prices of 1910–14) at a time when farm prices were very low not only absolutely but also relatively to other prices, or so they no doubt thought, and they naturally favored a national policy that promised better returns to them. But if the indexes upon which such a national policy were based were highly inaccurate, then either of two results might ensue in a given situation. If the results favored the farmers they would doubtless accept the windfall as their due; but if the results were unfavorable they would at the least lose some of their enthusiasm for that particular form of national policy. The plea here is not against a farm policy either favorable or unfavorable to the farmers. It is only against a policy anchored to a highly erroneous measuring stick. If national policy is to be based on changing price levels than the nation cannot afford the luxury of badly measured price levels. Any solution of the problem of industrial disproportion in our economy based on erroneous measurement will turn out to be no solution at all and will have to be done over again. If the analysis of the previous six chapters is accepted,

then the difference in price levels between 1910–14 and the present almost defies measurement, partly because what we measure is subject to so much error, partly because the whole economic system of living has changed to such an extent since 1910–14 because of inventions, technical developments, the exhaustion of certain resources, the discovery of other resources, etc., that there is no realistic basis for comparison of the price levels of these two widely differing periods.

The Pros and Cons of Chain Series. The chain series is a product of individual links. First it is necessary to calculate P_{01}, P_{12}, P_{23}, etc., giving indexes linking each pair of adjacent periods. Then the links must be fused into a continuous chain by multiplication, so that, for example, $P_{03} = P_{01} \cdot P_{12} \cdot P_{23}$. The individual links can be constructed by using the very best formulas (Fisher's superlative list, our 3, 4, 5, and 6) and will then give a highly accurate measurement of the price (or quantity) change between each pair of adjacent periods. The items in the continuous chain, the P_{01}, P_{02}, $\cdots$, P_{0k}, are derived from these links, and the question at once arises whether the chain index thereby has more validity than, or has any preference over, the fixed-base index for a long series. In the discussion of the fixed-base index great emphasis was placed upon the increasingly unsatisfactory character of P_{0k} as the distance between 0 and k increases. Does the chaining of links in any way avoid the weaknesses of the fixed-base method of making these long-range comparisons? It is very important to emphasize that the two methods are on somewhat similar ground in this situation. The difficulties of the fixed-base index arise from the increasing differences between base- and given-year weights (the q_1 and q_0 for price indexes, the p_1 and p_0 for quantity indexes), the effect of which was here measured by D, and from the increasing proportion of unique commodities, the effect of which was measured by R. These differences have not been eliminated from the basic data but the construction of a chain index does treat them somewhat differently. The result is believed to be an improvement over fixed-base indexes for long-range comparisons but not a completely satisfactory measurement, for a satisfactory measurement is impossible to attain.

The improvement which a chain index effects over the fixed-base for any long-range comparison P_{0k} lies in the fact that the P_{0k} fixed-base is calculated from the data of periods 0 and k and loses all contact with the p's and q's of intervening years and with all the adjustments that they have produced in the economy, including all the commodities that have been lost on the way as well as all those that have been newly found in the same time. The P_{0k} chain system avoids these errors of omission, for it is compounded of the whole set of changes in the p's

and q's *through time* and it has retained its contact with the actual changes of these intervening periods.[1] Therefore the P_{0k} chain is more accurate in this sense than is the P_{0k} fixed base.

But, with all this advantage conceded, there is still an important defect in these long-range comparisons. It is that the basic changes in an economy, which cannot be reversed in time, tend to make comparisons of widely separated price levels or levels of production a meaningless hodge-podge. It is again the question: were prices in the War of 1812 higher or lower than in World War II? There is no realistic answer, and any measurement which purports to give one may be manipulating certain facts of observation but not offering interpretations that have much significance. Do not forget that the proposition here maintained is essentially that P_{01} or $P_{(100)(101)}$ may have great validity but that as the distance between the two periods of comparison becomes greater the measurement may be so defective as to be for all practical purposes useless. The upshot of this position is that there is one kind of information of great validity and of great importance that can be obtained from an index series and another that cannot be obtained and should not be sought. The valid and accurate information which the index series will give is a measurement of the *direction* of change from *period to period*. This, of course, is provided also by the individual links. The measurements of doubtful value are the actual levels attained by the index at great distances from the base and the comparisons of these levels for different periods. The idea under discussion here can be expressed very precisely in the term of mathematics. If we think of the continuous series on a graph, we can say that we have great confidence in the slope (direction of change) at any point of the graph but that we do not have the same confidence in any given ordinate (price or quantity level). The chain index gives us all that is validly obtainable in an index number series, namely, the best measurement of direction of change. Neither it nor the fixed-base index provides measurements of change over long periods in which we can have much confidence.

Does the Chain Index Have a Cumulative Error? There is one argument against the chain index that has had a prominent place in the literature. It is that the chain index draws progressively away from the fixed-base index as the series progresses and that therefore the chain index is guilty of cumulative error. Notice the two parts of this proposition: (1) Chain and fixed-base indexes diverge. On this point there is no disagreement. Sometimes the chain index is greater

[1] The literature recognizes this fact by emphasizing the advantage which the chain system has in additions to and substitutions in the commodity list.

than the other, sometimes less, but they diverge. (2) The chain index has a cumulative error. This cannot be a valid conclusion unless it is also true that the fixed-base index is correct. The latter is not true, and the analyses in this book have shown why it is not true. Furthermore, whenever chain and fixed-base indexes disagree we have every reason, based upon these earlier analyses, to accept the chain index as the more valid of the two (having less error as measured by D and R) if the time periods compared are not so far apart that we must discard all comparison as worthless.

Authorities Pro and Con the Chain Index. It is interesting to notice briefly the names of a few important students of index numbers and the position they take with respect to the chain system. Alfred Marshall's article in the *Contemporary Review* in 1887 is, of course, one of the great classical papers on index numbers. It is in this paper that we find the first mention and advocacy of a chain index number as the best means of making long-time comparisons. Edgeworth, acting as secretary for a committee of the British Association for the Advancement of Science, published his first "Memorandum" on index numbers, also in 1887 and in it accepted Marshall's proposal for these long-term comparisons.[1] Mitchell in *The Making and Using of Index Numbers*[2] finds three objections to the chain index: (1) errors are likely to accumulate, (2) interpretation of the final results is simpler in the fixed-base system, and (3) the chain tends to drift away from the fixed base. These *defects* would seem to condemn the chain system utterly in his judgment, but he ends up by recommending that both series be constructed if resources permit and that then the differences between the two be called to attention. Walsh mentions that with any formula which gives a biased error the errors will accumulate from period to period and that the fixed-base system is an attempt to avoid this result. He says, however, that the way to preclude this difficulty is to avoid using formulas with errors which cumulate and then the formulas can be used "in the proper way, in the 'chain' system." Thus Walsh is an out-and-out champion of the chain system. Fisher's position has varied. He says,[4] "The chain system is of little or no real use." He lists arguments in favor of it: (1) the comparison of adjacent periods is more exact, (2) the year-to-year lines of the chain curve have correct slope, which is not quite true of the fixed base, and (3) the withdrawal and entry of commodities are easier. But he no longer

[1] Edgeworth, *Papers Relating to Political Economy*, Vol. 1, p. 218.

[2] Bull. 284, p. 86.

[3] See Walsh, *The Problem of Estimation*, pp. 84–85, 105.

[4] See Fisher, *The Making of Index Numbers*, pp. 308–312.

gives the same credit to argument 1 that he formerly did. The second point he thinks is so unimportant that for the better formulas the eye cannot distinguish between the two results on a curve (minute differences show up in the figures). He ends by saying, "the fixed-base system ··· is slightly to be preferred to the chain, because,

(1) it is simpler to conceive and to calculate ···
(2) it has no cumulative error, as does the chain system ···
(3) graphically it is indistinguishable from the chain system."[1]

Fisher's position is understandable when one remembers one very important factor in all his scientific work. To him science was of no consequence unless it could be used to improve practice, and he was frequently willing to make concessions on theoretical points in order to facilitate the introduction of improved methods. His position on points 1 and 3 in the quotation above can very well belong in this category. His position on point 2 is an unqualified error.

Fixed Base vs. Chain Indexes for Place Comparisons. This discussion of fixed-base vs. chain indexes has dealt almost wholly with index number series progressing through time. And this restriction is logical since most of the applications do apply to time sequences. But attention has been called from time to time to the fact that there is real interest in place comparisons for either price or quantity indexes. It is true that place comparisons are not strictly analogous to time comparisons since continuous movement through space, in the sense of a price index, does not seem to have the same validity as continuous movement through time. But the question of differences in price levels or in production levels between different areas is often raised, and there is some evidence of an increase in interest in this sort of measurement. The example referred to on an earlier page of the desire of the Ford Motor Company to set wages so that all Ford wages everywhere would provide a common standard of comfort to all is a case in point.

There is no reason why the previous analysis cannot apply to place comparisons as well as to time comparisons, and in this new application the errors indicated by D and R will again appear in their appropriate situations to confound the results of such comparisons. The principle still applies of high accuracy for short-range comparisons. The Ford company should therefore be able to measure the price level of Toledo or Cleveland upon the Detroit base with much confidence in the result. Perhaps the price level for New York or for Los Angeles as well. The point at which the analogy for time series will not stand up will be

[1] *Ibid.*, pp. 311–312.

in an attempt to compare Detroit and Los Angeles by chaining together a series of intermediate steps on the area basis, Detroit to Omaha to Denver to Salt Lake City to San Francisco to Los Angeles. In any use of index numbers on an area basis we are always limited to a single two-way comparison. There will be no series. The two-way comparison can be made, subject to the type of limitations indicated above and analyzed in earlier pages. When quantities (q) vary too much, for a price index, the D measure of inconsistency will be high. And when the list of commodities consumed in the two areas shows great variability, that form of error indicated by R, the indicator of homogeneity error, will enter our measurements. Subject to these limitations, index numbers may be constructed for place comparisons upon the same principles as for time comparisons.

One sort of place comparison should give us pause. The comparison that attempts to cross international borders will run into the problem of national variations in consumption and production habits based upon differences in resources and upon a whole host of human differences that have their main roots in the historic past. There is no certainty in this writer's mind that a price index for Warsaw which is constructed upon the Detroit base will have enough validity to justify its calculation. This part of Ford's problem is insoluble by the index number route, and he should have sought another way of meeting his problem of a proper wage for his Polish workers in Warsaw. If his workers in Warsaw had been Americans, the problem, though still complicated, would not have been so hopeless.

4. The Problem of Monthly Indexes

Whether the length of the time interval for which index numbers are to be constructed presents any problem is an issue that has not been faced yet. The argument has always been presented in terms of periods or years. It happens that the length of the period is a matter of importance in some instances; for example, a monthly index presents some problems that are not met in dealing with annual indexes. It is the seasonal factor that brings into play some influences which operate differently in periods of months than in periods of years. Consider, for instance, a production index that contains as one of its constituent series the figures of iron ore shipments on the Great Lakes. For a yearly index there is no difficulty in including the annual tonnage figures and obtaining the appropriate price as weight. But, for a monthly index, production figures fall to zero during months when the lakes are frozen over, and one at once faces the problem of *homogeneity*. Ore

shipments become a unique commodity for the winter months and must be omitted from the index. In calling attention to the complication that thus exists in calculating monthly indexes (and weekly indexes would only increase the magnitude of the problem), the author does not propose to engage in an extended discussion of the problem of seasonal movements in indexes or of their measurement. A single suggestion is that a basic series of annual indexes be constructed and that a scheme of splicing monthly indexes to them be worked out. Some work was done in the mid-1920's by Professor Warren Waite and Mrs. Dorothea Kittredge, of the University of Minnesota, on the problem of the proper orientation of monthly with annual indexes The subject still offers a fertile field for further work.

Part II • Current Construction Methods and Their Shortcomings

CHAPTER 8

The U. S. Bureau of Labor Statistics Index Number of Wholesale Prices

8.1. INTRODUCTION

The decisions of the preceding pages regarding the best way to measure a change in price or quantity levels between two periods, or to construct an index number series, do not conform in any full sense with actual practice in the construction of any well-known or widely used index published in this country. A brief review of the history of a few important indexes may therefore be worth-while for several reasons. First, an index number that has been published over a long period of time should show a development of method. For if it is widely used it will have the benefit of years of effort on the part of its makers to improve it and at the same time will have been able to profit from the criticisms normally expected from its users and from those serious students of index numbers who have studied it throughout its history. A second advantage of such a review is to bring into juxtaposition cases of actual practice and the proposals that have filled the earlier pages of this book. It is hoped that this will help to clarify the author's meanings and to show the bearing of his recommendations on actual construction methods. Part II, therefore, will examine two well-known and widely used index numbers, will give a short history of the development of each one, and will utilize the proposals of the earlier chapters as a basis of objective criticism of these indexes and as a source of recommendations for improvement if any such seem called for. The indexes that will be examined are published by the federal government. They are:

1. The U. S. Bureau of Labor Statistics index of wholesale prices.
2. The U. S. Bureau of Labor Statistics index of consumers' prices.

8.2. PERIODS IN THE DEVELOPMENT OF THE WHOLESALE PRICE INDEX

The wholesale price index number published by the U. S. Bureau of Labor Statistics is one of the oldest indexes in the United States, and it is without doubt one of the most important and one of the most widely used indexes. Its history to date covers well over a half century of development.

1. Forerunners of the Index

The first step taken in the United States toward the accumulation of a historical record of prices and their reduction to an index number came at the time of one of the most significant political developments in our history. The quarter century following the close of the Civil War was characterized by a long decline in prices and wages with a consequent growth of unrest among wage earners and farmers alike The period was marked by the growth of trade unions, the organization of the Granger movement among the farmers, the rise of the Populist party, and the differences over monetary policy between Democrats and Republicans.

In the early 1890's the U S. Senate appointed a committee with Senator Aldrich, Republican from Rhode Island, as chairman Senator Aldrich was at that time one of the leading Republicans in the Senate and doubtless exercised a great influence upon Congressional policy This committee, therefore, was important. It was known as the Committee on Wages, Prices, and Transportation, and its purpose was, in large part, to find an answer to the question whether wages or prices had fallen further since the close of the Civil War. The committee chose Professor Roland Falkner, of the University of Pennsylvania, to perform the technical job of supervising the collection of prices and the construction of a price index, the particular feature of the work of the Aldrich committee in which our interest centers. Falkner and his aids succeeded in gathering a body of data on wholesale prices going back to 1840. Prices were obtained from the books of merchants and manufacturers and from trade journals, and average annual prices were calculated. In the end there were 90 price series from 1840 to 1859 and 223 series from 1860 to 1891. From these figures relatives were calculated upon the 1860 base, and index numbers were constructed as averages of these relatives. Falkner experimented with weighted averages, using as weights some rather scanty data on family expenditures; but his most important index over this period was the

one constructed as a simple average of relatives. For this formula he had a precedent in the *London Economist* index, first published in 1869, and another in the Sauerbeck index number, first published in London in 1886.

This Senate report was the start in the collection of what is probably today the largest single body of price data in existence, the files of the U. S. Bureau of Labor Statistics. The second important step in the collection of price data and in the construction of price index numbers occurred when the U. S. Department of Labor at the close of the decade of the 1890's asked Professor Falkner to bring the Senate committee index up to date. His report in Bulletin 27 (1900) of the U. S. Department of Labor presented price data at quarterly intervals from January 1890 to July 1899 and indexes by groups and for all commodities combined for the same intervals. His published indexes were simple averages of relatives. The report published 142 series of actual prices and 99 series of price relatives for the period. The base period covered nine quarterly price dates, January 1890 to January 1892 inclusive, thus centering on January 1891, the exact terminus of the Senate committee series. Falkner made use of the personnel of the U. S. Department of Labor to gather his data.

2. Simple Average of Relatives Index, 1890–1913

Falkner's study was followed very shortly by the department's Bulletin 39 (1902), which is really the beginning of continuous periodic collection and publication of wholesale price data and of a wholesale price index. In this first report by the department, price data are recorded *by months* from 1890 to 1901 inclusive, but group indexes and an all-commodities index were constructed by years only. These were constructed as simple averages of price relatives with a broadened base covering the years 1890–99. There were, in all, 251 price relatives covering the entire period and an additional ten covering part of the period. With the publication of this report periodical price collection and the construction of a continuous series of wholesale price indexes became a permanent feature of the work of the U. S. Department of Labor. This work was shortly taken over by the newly organized U. S. Bureau of Labor Statistics.

From 1902 to 1913 inclusive, the basic methods of construction of the price index remained as they had been worked out by the U. S. Department of Labor in its 1902 report; that is, the index continued to be constructed as a simple average of relatives upon the 1890–99 base. One new problem arose when in 1908 it became necessary to drop eleven price series, and eleven new series (not substitutions for

those dropped) were added. It was the problem of how to introduce new commodities into the index. The procedure adopted involved two steps. First, for the group into which the new commodity was to be introduced, a group index was calculated for 1908 on the 1907 base. It covered the whole list of both the old and the new commodities for these two years. It did not, of course, include the old commodities that had been dropped. It was then multiplied by the previous group index for 1907 on the 1890–99 base, a calculation using the old list of commodities. The result was interpreted as the group index for 1908 on the 1890–99 base. The same procedure applied to the overall index. It will be recognized, of course, that the bureau had seized upon the principle of the chain index at this point as a means of solving its problem of changes in the commodity list. This was a significant step for the future because the bureau has utilized the chain method as a basic part of its calculation technique until relatively recent times.

By the close of this period, 1913, the files of the U. S. Bureau of Labor Statistics contained something over 250 price series of monthly data extending back to 1890 which were being used in the construction of the wholesale price index. In fact, the 1913 report [1] indicates that the annual index from 1890–1913 shows the commodity list varying between 251 and 261 for all these years. Indexes were now available on both an annual and a monthly basis from 1890 and were published annually in the bureau's reports on wholesale prices. Furthermore, in an appendix the annual index was carried back to 1860 by utilizing the data of the Aldrich Senate committee for the years before 1890.

There is another item to be noted about this segment of the bureau's index number history. By the early years of this century several other index numbers of wholesale prices were being published, notably those by Dun and Bradstreet. Price levels were being carefully watched as they were measured by the several indexes, and it is not surprising that any differences in results should attract attention. Although it is true that all these indexes showed the same general movements of price levels, it was nevertheless a fact that differences among them existed. These differences aroused curiosity and eventually criticism. And nothing is more natural than that a government bureau should take note of criticism, some of which was directed against the bureau's technical procedures. Indeed, by this time there had come to be some doubt about the adequacy of an unweighted average of relatives. The time was ripe for overhauling methods and for finding better ones if they existed. In this situation the bureau invited Professor Wesley

[1] See Bull. 149, p. 11.

Mitchell to come to Washington to make a study of its index number and to recommend changes if the need for them were found.

3. The Period 1914–20—Weighted-Aggregative Index

Mitchell's study was published by the bureau in 1915 [1] but was completed in time to dominate the revision of the wholesale price index in the 1914 bureau report (Bulletin 181). Two changes of the greatest importance for the future of the index came directly as a result of the Mitchell study. Both were formula changes. Weights were introduced and the simple average of relatives formula was discarded in favor of the aggregative form, in which the index, the fixed-weight aggregative as we now know it, became the ratio of two aggregate values, the values of a basket of goods priced at the two sets of prices of base and given years. Mitchell in his final conclusions had laid special emphasis upon what he called "the weighted aggregate of actual prices" as "preferable for measuring average change in the amount of money required to buy goods." [2]

Having decided to introduce weighting into its index number formula, the bureau had to decide upon weights and then gather them from whatever source was available and finally put them into form for use in the index number. It was inevitable that the bureau's first task back in the earliest days of its work with index numbers should be that of organizing the periodic collection of price data. At that time the collection of quantity or weight data probably seemed both of secondary importance and very difficult Indeed much of the literature of index number theory, at least as applied in practice, was then devoted to explaining that proper selection of commodities was an indirect method of weighting and, so the implication went, a sufficient method.

By 1914, however, the procedures of price collection had been routinized, leaving only the problem of expanding the commodity list. Furthermore, the question of accuracy of the index number had been raised and the problem of weighting was forced upon the bureau's attention. Mitchell's influence was great and pointed to the necessity of weighting. Therefore, when the bureau sought quantity weights to measure the importance of prices of this period, it decided that the weight of a commodity should represent the *physical quantity of the commodity marketed for a given year,* and figures were gathered for the year 1909. This was the most recent period for which full information was available. Much of the weight data came from the census

[1] Bull. 284.
[2] *Ibid.,* p. 113.

of manufactures and the census was then taken only at five-year intervals. These weights remained in use until the next important revision of the index in 1921, and thus they firmly established the policy of this period of using a *fixed-weight* aggregative index.

Thus the bureau had a new formula with fixed weights based on 1909 quantities. The formula was $\Sigma p_k q_a / \Sigma p_0 q_a$, where a refers to 1909. A new base was then decided upon. The old indexes from the beginning, in 1902, had been based upon average prices for the years 1890–99. For the new indexes, first published in 1914 and using the fixed-weight aggregative formula, it was decided that the base period should be the last completed year, 1914. The bureau explained that "this change was made for the purpose first of utilizing the latest and most trustworthy price quotations as the base from which price fluctuations are to be measured and, second, to permit the addition of new articles to those formerly included in the index." [1] This statement points to the bureau's recognition of the need for revising and expanding the commodity list as rapidly as it became known that some commodities on the list had been superseded in market importance by others and that commodities new to the list deserved recognition. We saw how in 1908 a chaining method had been devised whereby new commodities could be introduced into the simple average-of-relatives index. The same principles could be and were applied here to the aggregative form of index in order to increase the commodity list.[2] By making the last completed year the base for each succeeding year new commodities could be introduced easily in accordance with need or their availability. In other words, if for each succeeding year indexes were calculated on the previous year as base, then the largest possible list of commodities could be utilized in the index for each pair of adjacent years and the latest year could then be chained to previously constructed indexes to provide continuity for the series. The bureau adhered to this plan in its bulletins for 1914, 1915, and 1916. Apparently it went even further than utilizing the principle of chaining to introduce new commodities since for each succeeding year it published a whole new set of index numbers with the new year as base (100 per cent position). Whether this involved a complete recalculation of all the indexes whenever the base was changed, or just a shift to the new base by simple division, it is not possible to determine from the published reports. The revised indexes published while this policy of yearly base revision was in effect were the following:

[1] Bull. 181, p. 5.
[2] *Ibid.*, p. 255.

Report Year	Base	Indexes Published Annual	Monthly
1914	1914	1890–1914	1913–14
1915	1915	1890–1915	1914–15
1916	1916	1890–1916	1913–16

Before the revision the bureau had been keeping up to date a continuous series of annual indexes since 1860 by linking the Falkner (Senate committee) series before 1890 to the bureau series after that date. This procedure was repeated in the 1914 report,[1] still utilizing the old bureau series. The revised indexes based upon the weighted-aggregative formula were not linked in with the Senate committee index during the years 1914–20.

After 1916, annual publication was interrupted by World War I and was not resumed until 1919. In that year a combined report was issued for the years 1917–19 inclusive, and for these years the base was again shifted, but this time to 1913, "in order to provide a prewar standard for measuring price change." [2] This was a new policy of dealing with the base. It was a shift from a progressively changing base to a fixed base and at once became the established policy of the bureau. There was undoubtedly a widespread interest at this time in comparing prewar, wartime, and postwar prices. But one wonders if part of the reason for this reversion to a fixed-base index did not lie in a certain amount of confusion among users of the index from finding a new 100 per cent position every year. In any event, by 1921, the fixed-base, fixed-weight aggregative formula was established policy. With seven years of experience with 1909 weights, and with general recognition that these weights could not indefinitely measure the importance of prices of the future, the bureau's next important move in revising the index was, logically, a revision of the weights.

4. The Revision of 1921—New Weights and New Commodities

The revision of 1921 changed the construction features of the wholesale price index in two important respects. The weights were revised and the commodity list was greatly increased. The bureau had expressed the view, when the aggregative index number was adopted in 1914, that weights should be revised "every ten years as new census information should become available." [3] For the years since 1914, as we have seen, weights were based upon quantities marketed in 1909.

[1] *Ibid.*, Appendix III, footnote

[2] See Bull. 269, p. 8, and Bull. 296, p. 2.

[3] See Bull. 320, p. 9.

New weights were now available for 1919 from the census of manufactures and from other sources upon which the bureau had previously drawn, and they were one important reason for the revision that took place in 1921. But the time was appropriate also to utilize a larger commodity list that was now available, and this list, from which the new index numbers were made, included 404 items as against 327 for the previous year.

The introduction of these two changes did not involve any alterations in basic policy. The index was still a fixed-base, fixed-weight aggregative with the same 1913 base as previously but with new weights and an enlarged list of commodities. All these factors remained constant throughout the years 1921 to 1926. This fact did indeed necessitate a recalculation of index numbers for past years so as to form a continuous series with the new indexes of the future. In the 1921 report the publication of the revised indexes began, new yearly indexes being given from 1890 and new monthly indexes from 1913. In 1923, the revision of the monthly figures was carried back to 1900 and, in a final special bulletin for this period published in 1928, back to 1890.

The features, then, which differentiate the period 1921–26 from its predecessor were two: (1) revision of weights, which was in line with the policy contemplated by the bureau when it introduced its weighted aggregative index in 1914, and (2) increase in the commodity list. Since weights, commodity list, and formula remained constant throughout the period, the index gained a certain stability both of form and content. But there was one parallel to earlier periods, namely, that indexes were recalculated for all past years and months back to 1890. This recalculation provided comparability of the entire series up to the latest date. It was a task that must have absorbed a sizable portion of time and effort and of the bureau's budget. Such may be the price to be paid for progress! The necessity of recalculation cannot be avoided when new commodities are added to the list and when, their prices having been obtained for past years, it has been decided to include them in the index. Whether recalculation due to change of form, base, or weights can be dispensed with may be another matter. This question will be considered at a later stage.

One further event occurred during this period which belongs to index number history even though it is not a part of the history of the bureau indexes. Alvin Hansen had published an annual index of prices for the years 1801–40 in the *Quarterly Proceedings of the American Statistical Association* in December 1915 but had not included his original data. In 1923 the bureau published his original price data as

well as his index.[1] Hansen had calculated three separate indexes, each an unweighted average of relatives on the 1825 base. His three indexes were the arithmetic and the geometric means and the median of the relatives. His data included 142 series of prices from Boston and New York markets. The significance of these indexes lies, not in any contribution which they make to index number theory, but rather in the fact that they were shortly after this time to be combined with the Senate committee index figures to provide the only long-time measure of the course of wholesale prices from the beginning of the nineteenth century to 1890, the date of the beginning of the bureau index.

5. The Period 1927–36—Changed Base, Variable Weights, New Commodities

Just as particular changes in construction practice sufficed to set off the years 1921–26 as a distinct period in the history of the wholesale price index, so again the following period, 1927–36, is characterized by changes in both form and content of the index number. These changes involve commodity list, base, and weights.

Expansion of the commodity list is to be expected, of course, in the growth of an index number of the scope of the wholesale price index, particularly when the index is a governmental product. The extra cost of any improvement, especially the improvement that must follow a broader commodity base, is a small factor when it is recognized that decisions which turn on the index may involve sums running into millions of dollars. Therefore, no reasonable cost can be spared to make the index as accurate as possible. The users would be the first to support such a policy. The index, for example, was written into many war contracts during the last war as a basis for adjustments of certain costs. The evidence from these historical notes that the bureau has been constantly on the alert to expand the commodity list, both to make it broader and to make it more representative of all wholesale markets, is one of the marks of its desire continually to improve the index. This period 1927–36 brought two important expansions of the commodity list. In the revision of 1927, the commodity list used in the index was increased to 550 from the previous 404, and in 1931 there was a further increase to 784 series. These changes, of course, meant two sets of recalculations. First, the 1926 commodity list was priced back to 1923, and revised indexes were calculated to include these new data. Then, with the advent of the 784-commodity list in 1931, there was a recalculation of all monthly indexes back through 1926.

[1] Bull. 367, Appendix F.

For the new indexes beginning in 1927, the base was shifted to 1926. The bureau at this time explained that the shift was made "in order that the latest and most reliable information may be utilized as the standard for measuring price changes. Also it has been increasingly apparent that the year 1913 is now too remote to furnish a satisfactory base for comparing price levels in recent years." [1] This statement is taken from a special report of September 1927 publishing the first set of revised index numbers. The bureau added a further statement of its reasons for the change of base in the report for the year 1927 when it said that the 1926 base "furnished the most dependable standard for measuring price changes. Moreover, taken as a whole, market conditions in 1926 were regarded as fairly close to normal for the post war period." [2] There runs through both these quotations a thought that has appeared before in the bureau's periodic reports on its index: that comparisons of widely separated periods are not so important nor so often wanted as nearby comparisons. When the base was shifted to 1914 from the previous 1890–99, attention was called to the fact that comparisons between 1913 and 1914, or between 1908 and 1912, were generally of more interest than comparisons of 1914 with a distant base. This most recent change of base to 1926 was therefore completely in line with a policy that had by this time become definitely fixed.

A drastic change in policy, however, marked the third important change in this revision. The new weight policy was a complete reversal of the previous position. The bureau was earlier quoted in its 1921 report as having stated that the shift from 1909 to 1919 weights "conforms to the plan contemplated by the bureau at the inception of its weighted index number system in 1914 of revising the weighting factors every ten years as new census information should become available." [3] By 1927 it would seem that the policy of fixed weights for the aggregative formula, with a revision at periodic intervals, say ten years, was established policy, for it had been first put in operation in 1914 and it had now been kept in force for a dozen years. But some different thoughts must have been stirring somewhere within the bureau, for when Ethelbert Stewart, the Commissioner of Labor Statistics, addressed the New York Section of the American Statistical Association on October 27, 1927,[4] on the subject of the new indexes, his statement concerning weight revision indicated a basic change in policy. He said,

[1] See Bull. 453, p. 1.

[2] See Bull. 473, p. 2.

[3] See Bull 320, p 9; see also this volume, p. 90.

[4] The revision of 1927 was in full swing during the previous summer, and the first results were published in September, Bull. 453.

". . . the present plan is not only to add to the index new commodities as they appear but to reweight with each succeeding census; that is to say, the fixed weighting period has been entirely abandoned both in theory and in practice, and the weights will be revised with each census period."[1] At this time the census of manufactures was being taken every two years, and so the commissioner's statement meant new weights every two years. Speculation naturally arises as to why this radical change in policy; and its closeness in time to one other very important event in index number history cannot have been wholly accidental. Fisher's most important book *The Making of Index Numbers* was published in 1923, and many sessions at the American Statistical Association meetings in the years just before, during, and just after this year were devoted to the subject of index numbers and afforded the proponents of various views ample opportunity to express themselves. There was much discussion pro and con the Ideal formula, and that, of course, involves both base- and given-year weights. Commissioner Stewart's speech may well have reflected a shift in policy that followed long and serious discussion, within the bureau, of these new or newly emphasized methods of weighting, and it may have reflected a willingness on the part of the bureau to experiment with the new methods as a possible means of improvement of the index. The record of this period in the history of the wholesale price index is evidence that the system of current weights, in so far as the data were available, was going to be given a trial. It will be much simpler for the reader if this record is put down in tabular form as shown on page 94.

With these changes in weights and with the numbers of commodities changing several times over the period from 1913 it is clear that the construction of the revised index involved a linking process between consecutive periods when the changes occurred. The linking was apparently first used with the bureau's index in 1908 when a considerable change occurred in the commodity list. It has been used repeatedly since the adoption of the aggregative formula in 1914. Thus, where a change of weights or of commodities occurred two aggregates were calculated for the period in question, one based upon the weights and commodity list that was common to the preceding year, the other upon weights and commodities common to the following year. This, of course, is a typical example of the principle of the chain index and it is a proper description of the bureau's procedure to say that this principle was used extensively during the 1927–36 period. This introduction of new weights every two years and the chaining necessitated by

[1] Address published in *Monthly Labor Rev.*, December 1927.

Weights Used in Revisions of 1927–36 *

Indexes for Years and Months of	Weights
1890–1912	(No information)
1913	Mean of 1909 and 1914 quantities
1914–19	Mean of 1914 and 1919 quantities
1919–21	Mean of 1919 and 1921 quantities
1921–23	Mean of 1921 and 1923 quantities
1923–29	Mean of 1923 and 1925 quantities
1930–31	Mean of 1925 and 1927 quantities
1932–33	Mean of 1927 and 1929 quantities
1934–36	Mean of 1929 and 1931 quantities

* All this information, except the last entry, is available in the bulletins of this period, numbers 472, 493, 521, 543, and 572. The last entry comes from Cutts' and Dennis' article on the 1937 revision, *J. Am. Statistical Assoc.*, December 1937. The tabular form is incorrect in one minor respect: for some commodities the weights are averages of three years, of which the two end ones are as given in this table, for example, 1923–25 rather than 1923 and 1925. To put these additional details in the table would only add confusion and would be of no service for the purpose to which the table is directed.

changing weights represents the most radical step in index number construction in the entire history of the index. It is indeed a momentous landmark in the history of index number construction.

It is to be stated finally and briefly that the wholesale price bulletins of this period provided revised indexes, both annual and monthly, extending back to 1890. Just what the method of revision was before 1913, if it was anything more than a simple shift of base (that is, of the 100 position) by division, is not explained in the reports. The bureau also published during this period a long series of annual indexes from 1801 to date by piecing together Hansen's index, the Senate committee index, and its own. Finally, it published a monthly index (with important gaps, mainly before 1749) from 1720 to 1789 that had been constructed by Warren and Pearson, of Cornell University. Neither of these two long series indexes have significance for this study of practice in index number measurement. They merely give a fair, general idea of changes in price levels over a past century. They are the only measures we have.

6. The Revision of 1937 and the Return to the System of Fixed Weights

The year 1937 marks the final dividing line before World War II between periods in the history of the wholesale price index. The previous period, 1927–36, has been described as a most unusual one in the history of the index because it marked a complete reversal of

accepted practice in the matter of weight revision. The reason was that it had substituted for a system of fixed weights subject to revision about every decade a plan whereby weights were to be changed every two years, or as frequently as new quantity data were available from the census of manufactures. This change was here interpreted as a move in the direction of the radical position of Ideal formula advocates that weights should always be kept up to date. Whether or not such ideas influenced the thinking inside the U. S. Bureau of Labor Statistics during this period, the events of 1937 indicated no such views among the policy makers. The bureau next came up with a revision that completely reversed the practice of the preceding decade and reverted to the fixed-weight procedure of the years before 1927. The newly revised index from 1937 on became a fixed-base, fixed-weight aggregative. The base was 1926, the weights, estimated average quantities marketed in 1929–30–31. The bureau personnel explained[1] that the new fixed-base, fixed-weight system of construction replaced the chain system that had been in operation since 1908.[2] It will be remembered that an especial virtue of the chain system was that it made for easy introduction of both new commodities and new weights in the index. The new index, as of 1937 and thereafter, was presumably to solve the problem of new weights by not allowing them except at intervals of a decade or so and then by recalculating past indexes—the procedure that had been used from 1914 to 1926.

As for the commodity list, it was not possible to freeze the list and admit no new commodities, for no policy in the entire experience of the U. S. Bureau of Labor Statistics with its wholesale price index is more fixed than that of constant revision of the commodity list as some series became obsolete and others became important in the markets. This process of expanding commodity lists has continued throughout the history of the index, and, as of August 1948, the index included over 850 separate price series. Now the question became how to replace the chain system with a fixed-base, fixed-weight system and still permit the introduction of changes in the commodity list. A procedure was devised to deal with straight cases of substitution, the details of which need not be developed here. There is no record of how the bureau met the problem of introducing new series, not substitutions, in the index without employing a chaining device. These changes bring us up to the year 1949 in the history of this index. An attempt at evaluation of this historical record will be made after the corresponding review of the other index is completed.

[1] See Cutts and Dennis in *J. Am. Statistical Assoc.*, December 1937.

[2] *Ibid.*, p. 663.

CHAPTER 9

The U. S. Bureau of Labor Statistics Index of Consumers' Prices

9.1. THE MANY USES FOR A CONSUMERS' PRICE INDEX

The index of consumers' prices published by the U. S. Bureau of Labor Statistics is the bureau's second major development in the index number field, the index number of wholesale prices being the first. In its full development as an all-round commodity index in the consumer prices field it is of much later date than the wholesale price index, since the latter was begun in 1902 whereas the consumers' price index as an established and continuing product of the bureau did not begin until 1921. Its construction features have been much influenced by the technical methods developed for the wholesale price index.

Probably no index currently published in the United States is more widely used than this one, or more eagerly watched for its latest indications of change. Its immediate predecessors, constructed during World War I, played a part in wage negotiations when wartime price changes were upon us, and the permanent index, established soon afterwards, has been an almost constant feature of large-scale wage adjustments since that day. The Little Steel formula of World War II recognized this index as a measure of increased consumer prices between January 1941 and May 1942, and the National War Labor Board established it as a basis for wage adjustments for all industries. The index has found other applications: (1) as a basis for determining real wages; (2) as a guide to general economic policy in matters such as price control or rent control; (3) in international comparisons; (4) in the adjustment of claims for relief, unemployment, or other compensation; and (5) as a general measure of the price level in long-term contracts.[1]

[1] See *The Consumers' Price Index,* Joint Committee on the Economic Report, 80th Congress, 2nd session.

9.2. FORERUNNERS OF THE INDEX

1. Food Cost Indexes before 1919

Studies of the cost of living of particular industrial groups or in particular localities in the United States antedated the construction of any indexes of consumer costs by many years. The sixth and seventh annual reports of the U. S. Commissioner of Labor for the years 1890 and 1891 contained such studies. In 1890 a survey of 3260 families was published for iron, steel, and cognate industries, and in 1891 a survey of 5284 families for the cotton, woolen, and glass industries. These early studies set a pattern The eighteenth annual report of the commissioner, for the years 1901–02, contained a study of 25,440 families located in the principal industrial centers of thirty-three states. Several less comprehensive local studies have been published in the *Monthly Labor Review*.[1] Although such studies are an essential forerunner of any general index of consumer prices, since they are the source of (1) the lists of commodities purchased by consumers and (2) the quantities of and expenditures upon these goods, these early studies were not followed by immediate construction of consumers' price indexes There was a beginning, however, when the bureau first introduced a food cost index in 1903.[2] This index was carried back to 1890 and was continued to the days of World War I.

2. Cost-of-Living Indexes during World War I

It was during World War I that the demand for a more comprehensive index of consumer costs arose. Wage negotiations, especially in centers of rapid wartime expansion, such as shipbuilding areas, were complicated by unusual increases in retail prices One of the most forceful arguments of the workers for higher wages was that the extra costs should be covered in wages, or else the workers would be sacrificing for the war effort in terms of lowered standards of living and then presumably lowered efficiency. Several cost-of-living indexes, as they were then called, were constructed during the period to deal with such immediate problems.

3. The 1917–19 Survey of Family Expenditures

These wartime indexes produced two results of immediate consequence. First, they firmly established a precedent of wage adjustments based upon changes in consumer costs. The rule was applied, to be sure, in a period of rising prices, and it could be expected that the

[1] See references in Bull. 357, p 1, note.
[2] Bull. 54, pp. 1129–1163.

workers would put on the pressure for such adjustments only during rising prices. The logic of this sort of adjustment once accepted, however, the next important step was to establish regular and continuous indexes of costs of living, and this required comprehensive and up-to-date studies of family expenditures. The surveys of 1890–91 and of 1901–02 were too out-of-date to be applicable to the postwar conditions of the early 1920's. Thus the bureau conducted a survey of expenditures among workingmen's families during the years 1917–19.

This survey was the foundation for the first permanent index of cost of living, or of consumers' prices, constructed by the U. S. Bureau of Labor Statistics. It covered 12,096 white families in 92 cities and small towns in 42 states. For the total group the average number of persons per family was 4.9, the average yearly expenditure per family was $1434, and the average yearly income was $1513.[1] These averages were significant for purposes of index number construction, since the commodity list and, more important, the quantities to be used as weights, must of necessity be based on averages. Since for a given group only one index was constructed, it could not in any precise way apply to all the necessarily diverse conditions among all the families. It had to apply to a family *type* for the group, and this type was invariably determined by some sort of average. Criticism of the bureau's index during World War II was in some respects related to this matter of the difference between lower income levels within a group and the group average, as will be noted later.

9.3. THE ORIGINAL U. S. BUREAU OF LABOR STATISTICS COST-OF-LIVING INDEX, 1921–34

The 1917–19 survey of family expenditures was not so much a preliminary to the publication of a regular and continuous index of the cost of living as it was an integral part of the process. After the wartime experience of adjusting wages on the basis of changing consumer prices it was almost inevitable that this procedure would be continued in peacetime. The 1917–19 survey was a natural expansion of the work that had begun as special local surveys of shipbuilding centers, and, at the same time that the U. S. Bureau of Labor Statistics field staffs were collecting family expenditure data, they were also obtaining yearly price data. In some cities prices were gathered for December 1917 only; in others, for December each year from 1914 to 1917. This whole body of data became the foundation of the first regularly pub-

[1] Bull. 357, pp. 1, 4.

lished index of living costs issued by the U. S. Bureau of Labor Statistics.

1. Pattern Set by the Bureau's First Cost-of-Living Index

In some ways this first regularly published index set a pattern that has continued without modification to the present; that is, certain features of both coverage and measurement in the first index are as characteristic of it today as in the beginning. In content and in precision of measurement important changes have taken place, but not in form. The consistency of form is due partly to the status of index number theory in 1919 and partly to the pressure of problems that brought forth cost-of-living index numbers during World War I.

In the first place this index number has always been of the weighted aggregative type. That this form was selected for the first of these indexes in 1917–19 is unquestionably due to the influence of Mitchell and to his 1915 study. For the immediate result of that study[1] was that the bureau discarded a simple average-of-relatives form of index number in favor of a weighted aggregative for its wholesale price index. The logic of Mitchell's study was not, however, in any sense limited to particular kinds of price indexes; it applied to all of them. The result was that the weighted aggregative form became generally accepted as practically the best. Mitchell's prestige became very great, and any deviation from this newly established channel of thought on index numbers would have had heavy going. Furthermore, and this factor is always important, Mitchell's bulletin had the widest possible distribution among students of index numbers. Being a government bulletin and free for the asking, it had a vogue paralleled today only by a book-of-the-month selection. This is not to say that Mitchell's bulletin did not deserve the attention which it received. It did. But the post-script is important: No matter how significant the original discovery, if it does not become known, its usefulness is limited. Mitchell's advocacy of the weighted aggregative form was sound index number theory; but its great influence was partly due to the wide distribution afforded through publication as a government bulletin. The cost-of-living index originated in the same bureau that had put Mitchell's proposals into practice in the construction of its wholesale price index. The bureau naturally introduced the same advanced practices into the new index.

The other feature of the cost-of-living index for which a permanent pattern was established in the beginning was the coverage: Whose cost

[1] See p. 87.

of living did it measure? The index has been from the start a measure of the cost of living of the lower-income groups in a few large cities. As is generally stated in the bureau literature, the index is designed to cover wage earners and moderate-income clerical workers in large cities. The lower-income groups, but not the lowest, were those over which controversy arose during the period of high wartime prices, for it was typically those wage earners or lower-salaried groups in the centers of wartime industrial expansion who created the problem of wage adjustment. The concentration of attention upon the larger cities, of chief industrial import, is clearly evident in the area for which the bureau constructed its first indexes after the 1917–19 survey. Indexes were published for thirty-two cities. This feature of the index has not changed in character up to the present, although two cities have been added.

The original index, then, began as a weighted aggregative index applying to moderate-income families in the most important industrial areas and large cities in the United States. And it has continued to be this kind of index throughout the first three decades of its existence. The modifications that have occurred, and to which the next few pages will be devoted, have been concerned with some expansion of coverage in terms of cities, chiefly of temporary character during World War II, but mainly with the development of a greater sharpness of outline in the measurement itself, particularly as ideas have been clarified as to what a weighted aggregative should measure.

2. The Original Group and All-Item Indexes

The first publication of a cost-of-living index that was to be maintained thereafter as a permanent and recurring feature of the bureau's work occurred in 1921. The index was extended backward at that time for yearly intervals sometimes to 1917 and sometimes to 1914 and, following the original publication date, appeared currently at quarterly intervals. The basic indexes published on those dates were the city group indexes. At the start total family expenditures were classified into six groups and subindexes for each group along with a composite index were constructed for each of thirty-two cities. Indexes for the same six groups and the same thirty-two cities are still being constructed today. Two additional cities are now included. The accompanying tabular scheme indicates the relationship between the basic indexes (city group indexes) and the composite indexes, for each city and for the United States, that were constructed from the basic indexes. It will at the same time give a sort of reference guide to the discussion of construction methods that follow.

City Group Indexes and the Composites Obtained from Them

Groups	Cities by Number				U. S.
	1	2	...	32	
Food				→	→
Clothing					
Rents					
Fuel and light					
House furnishings					
Miscellaneous					
All items	↓	↓	↓	↓	↓

As said above, the basic indexes were the city group indexes, those which are comprehended within the body of the table. The arrows indicate the source of the several composite indexes: (1) group indexes for the United States, nowadays referred to as the national indexes, were constructed as combinations of the corresponding city indexes for the same group; and (2) all-item indexes for each city and for the United States were constructed as combinations of the several group indexes belonging to that city or to the United States. The construction methods, therefore, have three parts:

1. The construction of city group indexes.
2. The construction of United States group indexes.
3. The construction of all-item indexes, both for cities and for the United States.

3. Construction Details

(1) City Group Indexes. The city group indexes were constructed as constant-weighted aggregatives. In form they may be written

$$G_{0k} = \frac{\Sigma p_k q_a}{\Sigma p_0 q_a}$$

where G indicates a group index, and additional designation will be needed whenever it is necessary to distinguish between a city group index and a national group index. The subscripts 0 and k and a have the meaning that has been attached to them throughout this book. The year 1913 was selected as base, and the quantity weights, q_a, came

from the 1917–19 survey of family expenditures. Once the series of value aggregates $\Sigma p_k q_a$ were obtained, the remaining step in the calculation of group indexes was completed by simple division of the base-period aggregate into all others. In extending the first published indexes to dates prior to 1921, the bureau ran into a typical difficulty: the food index before 1920 contained only twenty-two foods, whereas thereafter it covered forty-two foods. A continuous series was made by a linking process on the December 1920 index, as follows: The December 1920 index for twenty-two foods was calculated on the 1917 base. Call it $P_{D\ 20}$. Then, after the calculation of $\Sigma p_k q_a$ for December 1920 and periods thereafter for forty-two foods, the formula for any period k of the continuous-series index on the 1913 base became

$$P_k = P_{D\ 20} \cdot \frac{\Sigma p_k q_a}{\Sigma p_{D\ 20} q_a}$$

This now-familiar linking process has apparently been used by the bureau whenever a change has occurred in the commodity list or weights.

The heart of the whole construction procedure at this point lies in the calculation of the terms $\Sigma p_k q_a$ (for $k = 0, 1, \cdots$). The commodity list that would appropriately represent the economic group whose living costs were being measured was obtained from the 1917–19 survey. This survey showed, in common with all others, that family consumption expenditures for any considerable group of families (there were 12,096 families in the survey) range over hundreds of different commodities in terms of kind or quality. Not all these commodities, of course, should be included in an index, first, because they would make the calculations extremely heavy and tedious, and, second, because a complete list is unnecessary. The principles of sample selection are, of course, better understood today than in 1919, but the sample list of commodities used at that time showed in a very real sense an appreciation of how to make such a list properly representative of the whole range of family expenditure for the purpose of measuring price change. The very division of the sample commodity list into the six groups given above is evidence. The first commodity list contained 165 different goods and services, in addition to rents.[1]

The prices that entered into these first city group aggregates were average city prices. The quantities were the *average actual* quantities of *each individual commodity or service purchased* as shown in the 1917–19 survey. But it was not necessarily the purchases in the indi-

[1] See Bull. 699, p. 15.

vidual city that were used for the city group indexes. In the food group indexes, for example, the q_a's for a given city represented quantities purchased in the region in which the city was located. There were five of these regions: North and South Atlantic, North and South Central, and Western. These two peculiarities in the determination of the q_a's for the city group indexes deserve special notice, for they illustrate the point, to which reference was made above, that the earliest cost-of-living indexes suffered, in comparison with later ones, in terms of precision of measurement. Had the principle of sampling representation been thoroughly understood in the beginning, each commodity used in the index would have been looked upon as representing, not only itself, but also in many cases other commodities in the total family expenditure budget. Thus the sum of all the weights when properly determined would have measured, not the significance of the 165 commodities in the index, but rather the significance of *all the commodities in the family budget.* The principle is clear: a particular price is chosen to represent a set of commodities, none of which except the one priced is in the index list. The weight of this commodity should therefore measure the importance, not of this commodity in the family budget, but of the whole set that it is designed to represent.

The second feature of the q_a's for city group indexes, namely, that they were regional quantities, is wrong in theory, or at least is not so accurate as city q_a's, unless there can be reasonable certainty that conditions bearing on p's and q's are fairly stable within a region. It is probable that the bureau in these early days shied away from *city quantity* weights through fear that a city sample might be too small. But a properly selected sample of from 80 to 200 families would not greatly worry a modern sampling specialist. It is a fact that both these features of the original quantity weights for city indexes have now been changed.

(2) U. S. Group Indexes from City Group Indexes. The first city group indexes, then, were ratios of two value aggregates $\Sigma p_k q_a$ with the p's and q's as defined above. Group indexes for commodities other than foods for the United States were constructed from these thirty-two city group indexes. The U. S. food index was obtained by an extension of the procedure for calculating a city food index. $\Sigma p_k q_a$ was again calculated, but in this instance p_k was the average price of each food item in fifty-one cities from which food prices were obtained. The q_a's were the quantities purchased by 1917–19 families as measured in the entire survey. The index was, as usual, the quotient of the base-period aggregate divided into each of the other aggregates in turn. For

groups other than food the city aggregates were combined by simple addition, thus:

$$\sum^{32} \sum p_k q_a$$

where $\Sigma p_k q_a$ is as defined earlier for a given city and the first summation sign above indicates a summation of such aggregates for thirty-two cities. Given these combined aggregates, the usual process of division by the base magnitudes gave the indexes desired.

(3) All-Item Indexes for Each City and for the United States. The all-item indexes for each city and for the United States were constructed as weighted averages of the respective group indexes of each locality involved. The weights used were the *proportional* expenditure of the particular group as measured for that locality in the 1917–19 survey; that is, for each all-item city index the proportional expenditure weights came from *that city's* expenditure figures in the 1917–19 survey. For the all-item United States index the weights came from the entire 12,096 families in the survey.[1]

It may help to see this procedure in its entirety if it is put in formula form using present terminology. Let $G_1 \cdots G_6$ be the six group indexes for a specified city. Each will be of the form

$$G_i = \frac{\sum^{N_{0k}} p_k q_a}{\sum^{N_{0k}} p_0 q_a}$$

Then the all-item index P_k for this city will be

$$P_k = \sum_{i=1}^{6} G_i \left(\frac{\sum^{N_{0k}} p_a q_a}{\sum_{i=1}^{6} \sum^{N_{0k}} p_a q_a} \right) = \sum_{i=1}^{6} \left[\frac{\sum^{N_{0k}} p_k q_a}{\sum^{N_{0k}} p_0 q_a} \left(\frac{\sum^{N_{0k}} p_a q_a}{\sum_{i=1}^{6} \sum^{N_{0k}} p_a q_a} \right) \right]$$

It will be readily seen from this form that the all-item city index (or United States index) does not reduce to a simple aggregative form unless year 0 and year a are identical, which in this instance they are not.

9.4. THE 1935 REVISIONS—TECHNICAL IMPROVEMENTS[2]

The year 1921 thus saw the establishment of cost-of-living indexes for moderate-income groups in large cities of the United States, with comparative figures for the past, extending in some instances to 1914,

[1] See Bull. 357, pp. 456–466.

[2] See *Monthly Labor Rev.*, September 1935, pp. 819–837.

and with the prospect that new indexes would be compiled currently in order to keep a continuous index and to maintain comparability between present and past. The indexes were formed essentially on the principle of the constant-weighted aggregative with weights taken from the 1917–19 family expenditure survey. In this form the index continued without change until 1935.

1. The Need for Periodic Weight Revision Recognized

But constant-weighted aggregatives, even as presented in the Mitchell study of years earlier, were not a fixed procedure to be maintained for all time. It was recognized when the constant-weighted aggregative form was first advocated and when it was first introduced by the bureau in its wholesale price index in 1914 that the quantity weights must maintain a reasonable relationship to the prices that were being weighted. Thus it was early suggested that weights should be revised every ten years or so. This position can easily be accepted as an essential part of the inheritance that the cost-of-living index received from its predecessors. Its acceptance is evidenced in the fact that the bureau was unwilling to establish a permanent cost-of-living index after World War I based upon weights then available from earlier family expenditure surveys. The only survey of comprehensive character that had been made was the one for the years 1901–02. It had covered 25,440 workingmen's families in thirty-three states. A new survey was obviously a necessary part of a new cost-of-living index, and the 1917–19 survey of 12,096 families was planned and executed with the definite purpose of providing commodity lists and weighting factors for this index. Indeed, many of the price data for the new index, back as far as 1914, were obtained at the same time and apparently by the same field crews that gathered the family expenditure data.[1]

In conformity with the general theory of revising weights after a reasonable interval, it is as to be expected that the bureau had plans for new family expenditure studies in the not-too-distant future after the index was first established. These surveys, of course, served other purposes besides providing commodity lists and weights for price indexes. Three such surveys in the period 1921 to 1935 deserve notice. They are:

1. Expenditures of federal employees in five cities, 1927–28.
2. Expenditures of Ford employees in Detroit, 1929.

[1] See Bull. 357, p. 72.

3. Expenditures of federal employees in the District of Columbia, 1933.[1]

These surveys, as is evident, were local in character and were not designed with a view to revision of weights for the cost-of-living index. The Ford Motor Company study, for example, arose out of Ford's desire to establish common living standards, and wages based upon them, for his employees the world over. There can be little doubt, however, that the U. S. Bureau of Labor Statistics carried out these local surveys having in mind the possibility that the time would come when another comprehensive family budget study must be made and that its consumers' price index must be revised on the basis of it. Indeed, the plans for such a survey were completed and funds provided so that the work began in the summer of 1934. The budget study of 1934–36 was the result and was the basis for general revisions first published in 1939.

2. The 1935 Changes Were Methodological Improvements

The year 1935 marks the date of the first modifications in the form of the index after its first publication in 1921, and the years 1934–36 were the period during which the new comprehensive expenditure survey was being conducted. These two events were related only in time, however, not in causal sequence. Comprehensive revisions were planned to be made as soon as possible after the 1934–36 expenditure data became available. But the 1935 changes were not a matter of new content or new data. They were, rather, new ways of handling any body of price and quantity data such as had been utilized in the original index. They gave new form to the index, and they doubtless arose from the bureau's experience in studying its methodology during the one and one-half decades since the index was established. The bureau must have felt under pressure to introduce at the earliest opportunity any important methodological improvements that could be made in the construction of the index. It made four important changes in its procedures at this time. They concerned the weighting systems for:

1. The city food indexes (two changes).
2. The United States group indexes
3. The all-item city indexes

[1] Made jointly by the U. S Bureau of Labor Statistics and the Bureau of Home Economics of the U S Department of Agriculture.

3. Technical Changes

(1) Revision of the City Food Indexes. The city food indexes were expanded at this time to include eighty-four commodities in place of the forty-two that had been carried since 1921. This increase in coverage was important in terms of representation of the food budget but was in addition to the two methodological changes listed above. The latter were, first, the replacement of *regional* quantities as weights for city prices by city quantities and, second, the introduction of what the bureau called imputed weights.

It will be recalled [1] that the quantities (q_a) used in the original city indexes were, for each commodity, *the actual quantity of that good purchased,* according to the 1917–19 survey, *in the region in which the particular city was located.* The entire list of cities in the survey was divided into five groups. To illustrate, the North Atlantic region contained cities as widely separated, both physically and economically, as Portland, Maine, Boston, Buffalo, Philadelphia, and Scranton.[2] The food prices in each of these five cities, however, were weighted with the same q_a, the survey figure on quantity purchased *in the North Atlantic region.* Furthermore, the q_a represented only the quantity of the particular commodity; there was no sampling representation to cover foods that were in workingmen's budgets but not in the index.

The significance of the weighting changes in the food index is readily apparent when the new procedures are contrasted with the above. The bureau had found, by study of the 1917–19 survey, that food consumption showed great diversity from city to city. "It was therefore decided that, in so far as adequate figures were available, revised weights should be computed for the food-cost index for each city in which prices are secured, *based on the food purchasing habits of that city.*"[3] The q_a's now became city quantities rather than regional; or, in a few instances in which city samples were small and several nearby cities were similar in their economic setup, the q_a's referred to purchases in two or more nearby cities. For example, in Boston the q_a referred to Boston alone, but in Bridgeport, Providence, and Fall River the q_a's were the purchases in these three cities and in Lawrence. It was clearly felt that this change gave quantity weights more properly representative of the particular cities than the previous regional quantities, and of course the modern view is in agreement.

[1] See pp. 102–103.
[2] See Bull. 396, p. 3.
[3] See *Monthly Labor Rev.*, September 1935, p. 821. Italics added.

The second weighting change introduced into the food indexes was even more significant and probably represented the greatest single advance in methodological thinking introduced at this time. It was the introduction of what the bureau called imputed weights. Whereas a q_a had previously represented the *quantity purchased of the particular good priced*, it now came to have a wider meaning. It became a *quantity* representative of the whole list of commodities (when there were more than one) that were represented in the index by this price. The change was of great significance for it recognized the importance of the sampling problem. An actual workingman's expenditures during a year may include as many as 1400–1500 individual commodities (a bureau estimate [1]) but the index is based on less than 200. Therefore, the commodities included in the index either represent themselves only and therefore but a part of the workingman's consumption, or they represent the total consumption. In the original weighting system for the food indexes it is unquestionable that the narrower view of representation prevailed. The problem of sampling representation was not so clearly understood. The change whereby q_a came to represent not only p but, where sample representation is needed, a whole list of p's of which only one appears in the index was a fundamental one and marked an important technological advance. The procedure whereby the revised q_a could be obtained from the 1917–19 expenditure survey data can be very easily indicated. Suppose that p is the price of carrots, on which the average family expenditure in city A in 1917–19 was $v_1 = p_a q_a$. Suppose also that the index commodity list contains no other *vegetable;* and let the average family expenditure on all vegetables, 1917–19, in city A be v_2. The latter is a sum of money, as is v_1, but cannot be expressed directly in terms of p, q, at least without a price index for all vegetables. We have then

p_k = price of carrots for some period (k)
$v_1 = p_a q_a$ = average amount in dollars spent per family on carrots in 1917–19 in city A
v_2 = average amount spent per family on all vegetables in 1917–19 in city A

Then two weights (q's) for p_k can be established as follows:

$$q_a = \frac{v_1}{p_a} = \text{the weight used in the original index}$$

[1] See *The Consumers' Price Index,* Joint Committee on the Economic Report, 80th Congress, 2nd session, p. 4.

$q_a' = \dfrac{v_2}{p_a}$ = a hypothetical quantity weight for p_a that gives p_a an *importance* based upon *all vegetables* and not based upon carrots alone

The introduction of weights q_a' in place of q_a was the most important change made in the index in 1935.

(2) A New Method of Calculating National Group Indexes. The revision of the method of calculating the national group indexes affords another example of the progress in thinking on technological procedures. The original national indexes had been constructed[1] by one procedure for foods and by a different one for other groups. For the food index, average prices p_k were determined from all quotations in fifty-one cities and a single cost aggregate $\Sigma p_k q_a$ was then constructed, where q_a represented the average quantity purchased per family among all families in these fifty-one cities in 1917–19.[2] Other group aggregate costs were obtained by simple addition of the corresponding aggregates $\Sigma p_k q_a$ from each of thirty-two cities. The national aggregate was then $\sum^{c}\sum p_k q_a$, where the first summation sign indicates summation over cities, the remainder of the expression referring to summation within each city. In the revision of the national group indexes an average index for the United States was formed by calculating a weighted average of the city indexes for the group, the weights being proportional to populations in the several cities. This was an average of fifty-one city indexes for foods and of thirty-two for the other groups.[3] This change also represented a methodological improvement because of the fact that the different city indexes or aggregates represented different numbers of families, and these differences were not reflected in the quantity weights, q_a, since the weights represented *average* purchases per family in each city. The shift to proportional population weights in combining city indexes into national therefore gave a measurement in an appropriate way to the varying importance of an average family q_a in the different cities, since numbers of families vary in close conformity with total population.

(3) A New System of Calculating All-Item City or United States Indexes. The final change made in 1935 concerned the construction of all-item indexes for each city and for the United States. The earlier procedure had been to construct a weighted average of the group indexes, the weights based upon proportional expenditures by groups in

[1] See pp. 103–104.

[2] *Monthly Labor Rev.*, September 1935, pp. 820–821.

[3] See *Monthly Labor Rev.*, September 1935, p. 828, and Table 8, p. 836.

each city, or in the United States, according to the 1917–19 survey. The weights by groups, therefore, were of the form

$$\frac{\Sigma p_a q_a}{\sum^{G} \sum p_a q_a}$$

and the important point to note here is that all subscripts of the p's and q's are a; that is, the percentage weights were based wholly on the 1917–19 expenditure data. Attention was called a few pages earlier [1] to the fact that the all-item indexes constructed in this manner did not reduce to simple aggregatives, the formula being

$$\sum^{G} \left[\frac{\Sigma p_k q_a}{\Sigma p_0 q_a} \left(\frac{\Sigma p_a q_a}{\sum^{G} \sum p_a q_a} \right) \right]$$

In 1931 Hogg called attention [2] to what she considered a distortion in the index caused by these weights Her argument was in substance that, since the bureau measured the importance of a price p_k by the quantity q_a, a set of group indexes $G_k = \Sigma p_k q_a / \Sigma p_0 q_a$ should be weighted by the proportionate cost of the quantities q_a priced at the base-year prices p_0. Her proposal reduces to proportional expenditure weights for the G_k as follows:

$$\frac{\Sigma p_0 q_a}{\sum^{G} \sum p_0 q_a}$$

When these weights are used the weighted G_k's become

$$P_k = \sum \left[\frac{\Sigma p_k q_a}{\cancel{\Sigma p_0 q_a}} \cdot \frac{\cancel{\Sigma p_0 q_a}}{\Sigma\Sigma p_0 q_a} \right] = \frac{\sum^{G} \sum p_k q_a}{\sum^{G} \sum p_0 q_a}$$

which reduces to a fixed-weight aggregative for the all-items index with q_a weights, or its equivalent, a value weighted average of price relatives using $(p_0 q_a)$ values as weights.

The above seems to represent the principle involved in Hogg's criticism and it also represents in its essentials the change introduced in 1935 into the calculation of all-item indexes. The bureau's actual procedure involved, first, the calculation of aggregate costs for each group for 1913, here called $\Sigma p_0 q_a$. This figure for each group was ob-

[1] See p 104

[2] In *J. Am Statistical Assoc*, Vol. 26, pp 52–57.

tained by dividing the 1917–19 average expenditure for the city (or for the United States) by the group index of prices for 1917–19 on the 1913 base. The latter was actually a simple average of the two indexes for December 1917 and December 1918. The process is illustrated for the United States figures [1] herewith:

Group	1917–19 Average Expenditure	1917–19 Price Index on 1913 Base	Quotient $\Sigma p_0 q_a$
Food	563.92	172.2	327.57
Clothing	233 26	181.3	128.68
Rents	204.66	102.7	199.25
Fuel and light	71.68	135.0	53.09
House furnishings	66.45	177.8	37.37
Miscellaneous	296.21	151.9	194.98
	1436.18	152.6	940.94

The figures of the last column were then multiplied by the group indexes for any date, producing a set of group dollar cost aggregates for that date which we will call $\Sigma p_k q_a$. These when summarized represented the aggregate cost for all items $\Sigma\Sigma p_k q_a$ and when divided by $\Sigma\Sigma p_0 q_a = 940.94$ gave the desired index. The process is illustrated below:

Group	1913 Costs $\Sigma p_0 q_a$	G_k December 1918 on 1913 Base	Product $\Sigma p_k q_a$
Food	327.57	187.3	613 55
Clothing	128.68	213.4	274.66
Rents	199.25	105.3	209.87
Fuel and light	53 09	146.0	77.49
House furnishings	37.37	205 0	76.62
Miscellaneous	194.98	163.3	318.48
Aggregate costs	940.94		1570.67
Indexes	100.00		166.9

The bureau comments that "this procedure is necessitated by the fact that the index must be computed currently by the link-relative method, because of constant changes in the form in which consumer goods are offered for sale." [2] In other words, for any month or year in which changes occur in the commodity list, the current index can be linked to the previous period by calculating group indexes $G_{(k-1)k}$ (where k is

[1] Figures taken from *Monthly Labor Rev.*, September 1935, p. 827.

[2] *Monthly Labor Rev.*, September 1935, p. 827. Figures above from same source.

the current period) and carrying out the multiplications upon $\Sigma p_{k-1}q_a$ in the same form as above.

This final change of the 1935 revisions is also a distinct methodological improvement. When, as originally, the percentage weights for the group indexes were based upon the 1917–19 expenditures they represented a magnitude that could vary in two ways from the conditions of the base and given periods of a particular index figure, since both price and quantity elements in the weights came from period a. In the revised system only the quantity factor, q_a, came from outside the two periods compared. The bureau puts the matter in slightly different terms. On the assumption of a fixed consumption schedule (the quantities of 1917–19) the proper weights to use in combining group indexes into an all-item index are the proportional distribution among groups of the base (1913) cost aggregates of the first consumption schedule employed. The cost aggregates here referred to are Σp_0q_a. Whether in taking this position the bureau was influenced by the fact that, with a constant commodity list, this method of weighting reduces to the equivalent of the constant-weighted aggregative $\Sigma p_kq_a/\Sigma p_0q_a$ only the bureau can tell. Hogg said,[1] "In considering this question of consumption (i.e., the q_a's) the price change element (i.e., p_a) needs to be eliminated · · ·." And again, "· · · until new budgetary data are available, the most desirable course seems to be to regard the 1917–19 data as typical of 1918 and to assume as usual that expenditure changes in earlier and later years were due to changes of price, not of consumption habits."[2] Her position is apparently accepted in toto by the bureau

9.5. THE NEW FAMILY EXPENDITURE SURVEY OF 1934–36 AND THE CONSUMERS' PRICE INDEX AFTER 1939

1. The Cumulative Evidence of Changed Consumption Habits

By the middle 1930's the budgetary records of the 1917–19 family expenditure survey were becoming ancient history in a very real sense. Evidence is plentiful for the view that a new survey was needed to provide more up-to-date information on consumption habits and that the bureau was prepared to make a new survey at the earliest possible date. When Hogg in 1931 proposed a new method of weighting group indexes to avoid what she called a *distortion* she did so with the clear proviso that it was a *best* procedure only until new and current budgetary figures could be obtained. She said, *"A new cost of living*

[1] *J. Am Statistical Assoc.*, Vol 26, pp 55, 56.

[2] *Monthly Labor Rev*, September 1935, pp 825–826.

inquiry is much needed, but until new budgetary data are available, the most desirable course appears to be · · · ."[1] Hinrichs, the acting Commissioner of Labor Statistics, stated in the preface to the summary volume of the 1934–36 expenditure survey, "It had been apparent for some time before the present survey was initiated that consumption habits had changed greatly since the last investigation of this kind."[2] Attention was earlier[3] called to several small-scale surveys that were made in the late twenties and early thirties and to the fact that funds were allocated in 1934 for a comprehensive family budget study, the one which now carries the date of 1934–36.

It is worth a momentary pause to let others state how consumption habits had changed between the two survey dates of 1917–19 and 1934–36, and this task cannot be better performed than was done by the bureau itself:

"It had been generally recognized for some time that there was need for the introduction of new items into the index. Consumption habits have changed greatly since 1919. In the period since the end of the last war, the purchases of wage earners and clerical workers in the United States have included a great variety of consumers' goods which were not available previously. *Some of these goods were actually new*—rayon fabrics, for example, and certain types of electrical equipment. *Some of them had been in the market before,* but at prices higher than the moderate income families could pay. *Some of the differences were merely changes in fashion and custom.*

"Isolated studies of expenditures had shown that many more wage earners and low-salaried workers were living in houses with electricity than had been the case at the end of the war, that many of them were buying automobiles and radios, some of them were buying electric refrigerators. Fashions in dress had changed so much that it became apparent that mere substitution of a new type of garment for the equivalent of one previously worn did not represent contemporary clothing purchases."[4]

This is an important quotation, for it states concisely the circumstances that change people's habits of consumption. The circumstances are in no sense a temporary phenomenon. Consumption habits change (1) because new commodities are created (progress in invention). They change (2) because prices change: prices may come down so

[1] *J. Am. Statistical Assoc.,* Vol. 26, p. 56.

[2] Bull. 638, p. VIII.

[3] See pp. 105–106.

[4] Bull. 699, p. 2. Italics added.

that some can buy who could not afford it before (more progress; if not new invention, at least technical advance and thus cheaper costs of production). Finally, they change (3) because people's tastes change (and this to most of us is still more progress). These three factors in price change are not independent; indeed, they are very much interdependent. Altogether they spell lack of stability in family expenditure patterns. *It may be legitimate to question whether, in the face of such progress,* a new budgetary study every fifteen or twenty years is enough. This type of question will occupy some of the attention of the next, and last, chapter of this book, the evaluation of current practices.

2. The 1934–36 Expenditure Survey

The comprehensive survey of 1934–36 gave details of all family expenditures for a total number of 14,469 families. A few overall comparisons between this and the 1917–19 study will serve to show both similarities and differences between the two and may point up some of the reasons why a new survey was required. The former survey covered 12,096 families. The two are fairly similar in this respect. Granted no essential advantage of one survey over the other in the distribution of coverage within the areas studied (and the point is not to be considered here), either sample is large enough to give satisfactory estimates of the details required to determine expenditure and quantity averages for an index number. The new survey covered both white and colored families, the latter group comprising 11 per cent of the total; the earlier investigation covered only white families.[1] This difference is probably more than a reflection of changed labor conditions. There were colored laborers in 1919. They were merely neglected in that survey. The early survey collected data in ninety-two cities and towns, some of the towns having a population of only a few thousands. To be sure, indexes were constructed for only thirty-two large cities. That circumstance plus better understanding of sampling techniques may in part explain the fact that the new survey was confined to forty-two cities each of over 50,000 population. The average size of a family in the earlier survey was 4.9 whereas in the later one it was 3.6. These figures were in no sense a result of different kinds of selectivity in taking samples. They undoubtedly reflect a basic change that has been taking place in American family life for many years and that has attracted attention far beyond the range of index number makers. It is curious and surprising to find that there

[1] See Bull. 357, p. 1.

is little difference in average family income for the two dates: $1513 in 1919 and $1534 in 1935. Most guesses would have placed the latter figure much higher than the former, barring the contingency of a large discount caused by the nearness of the 1935 date to the low point of the depression of the thirties.

One final comparison of the two surveys, showing for each the percentage distribution of actual family expenditures among the several commodity groups, gives some indication of the changed consumption pattern of the mid-1930's as compared with that which immediately followed World War I: [1]

THE PERCENTAGE DISTRIBUTION OF ACTUAL FAMILY DISBURSEMENTS IN TWO SURVEYS

Group	1917–19 Survey	1934–36 Survey
Food	37.5	34.0
Clothing	15.5	10.3
Rent	13.6	17.5
Fuel, electricity, and ice	4.8	6.7
House furnishings	4.4	4.0
Insurance and other savings	7.2	0.7
Miscellaneous	17.0	26.8
	100.0	100.0

On a hasty view the outstanding shifts here appear to be in clothing, possibly rents, savings, and miscellaneous. The savings item may be explained in terms of the temporary economic circumstances of the workers with full employment after World War I and with probably very limited resources after the depression days of the early 1930's. The great increase in the miscellaneous item is significant. In sum, this simple table argues sufficiently for frequent budgetary surveys, for it shows significant shifts in expenditure patterns.

3. The 1940 Revisions

New Content. Following policy that was well established by this time, the index revision that succeeded the completion of the new expenditure survey included a change of base, more frequent publication, a new commodity list, new weights, and at least minor changes in technical methods. The new base for revised indexes was set as the five-year period 1935–39. This was in response to a recommendation of the Central Statistical Board; and the recommendation was in

[1] Figures taken from Bull. 699, p. 24, Table VIII.

support of a concerted effort among many statistical bureaus and boards in Washington to introduce similar procedures where possible and thereby promote comparability. The particular five years were no doubt selected with the thought that they represented *normality,* and it was not long before this base was spoken of as a prewar normal. The consumers' price index had been published twice yearly from 1925 to 1935 and mostly quarterly thereafter. The new series was for quarterly periods from 1935 to September 1940, but after the latter date it was put upon a monthly publication basis.

The commodity list was of course revised, as was inevitable with data of the new family expenditure survey available, and the details of these modifications, along with new weights, constitute the important changes in *content,* as distinct from form, in this most recent revision. The accompanying figures constitute Table II, from the bureau's Bulletin 699,[1] describing the new indexes.

NUMBERS OF GOODS AND SERVICES IN OLD AND NEW INDEXES *

	Original Index		New Index
Group	1919	1939	1939
Food	42 †	84	54
Clothing	61	63	48
Fuel, electricity, and ice	6	6	10
House furnishings	21	16	26
Miscellaneous	35	33	60
All items	165	202	198

* Not including rents.
† According to the 1935 revision of the 1919 index.

Several significant changes in numbers of commodities are here evident in the new index, and some indicate that not all improvements come by way of *increase* in the commodity list. For the food list has been decreased from the eighty-four commodities that it had contained since the 1935 revision to fifty-four in the new index. The bureau speaks of an *experimental* increase at an earlier date and states that the reduction from eighty-four to fifty-four foods was possible because the prices of some foods could be predicted from those of others.[2] The modern way of expressing the same idea is to say that fifty-four foods selected in such a way as to be a good representation of the

[1] *Ibid.*, p. 15.
[2] Bull. 699, pp. 15–16.

whole food budget is a better sample than eighty-four without such careful selection; or that, with both on an equally good selection basis, the fifty-four may be preferred to the eighty-four out of considerations of the relative effort and costs associated with the two lists.

The commodity list for clothing has a similar explanation. Many of the clothing items in the original list have been dropped, mainly boys' and girls' clothing. There have been a few additions. It is to be emphasized here that the issue is the very important one of proper sampling representation of the whole family budget. It is a question of fact whether given prices can represent others on a sampling basis. If they can, then one price can always represent such a group of prices, subject only to the errors of sampling. But sampling errors are generally measurable in any practical situation, or at least they can be estimated with sufficient accuracy to satisfy practical needs.

The increased numbers of commodities found in the utilities, house furnishings, and miscellaneous groups clearly represent better sampling, at least of the current family budget, than did the old list. They include such items as fuel oil and ice, among utilities; radios, washing machines, vacuum cleaners, and refrigerators, among house furnishings; and an extended list of items in the miscellaneous group among which automobile costs and beauty shop expenses are prominent. Most of the items here named are *new* commodities for these income groups and therefore could be obtained only in a modern commodity list; they belong to the family expenditure survey of 1934–36 and did not exist in the survey of 1917–19.

Technical Methods. New commodity lists and new budget weights are the sources of improved content of the new indexes. The other element in an accurate measurement of price change is form or technical method. In other words, the right data, the right commodity lists, and the proper prices and quantities, and, then, the right formulas. Content and form together constitute the foundation of accurate index number measurement. The changes in technical methods were not great in the 1940 revision. It is probably not too inaccurate to say that they went no further than to clarify the general trend, that was clearly apparent in 1935, to put as many indexes as possible in terms of the ratios of two cost aggregates, thus making the real task of construction that of calculating these cost aggregates Σpq for each group and for each city and, finally, for the United States. The following is a brief statement of the procedures employed in the new indexes of 1940, with the notational terminology of this book introduced for clarification:

1. City Group Indexes. A cost for each commodity was first calculated for March 1935. Let it be called $p_k q_a$ (k = March 1935) Then

$$p_k q_a = \frac{p_k}{p_a} v_a = \frac{p_k}{p_a} (p_a q_a)$$

where p_a = average price of *this commodity* for the year covered in the family expenditure survey of 1934–36

p_k = average price of *this commodity* for March 1935

v_a = average family expenditures for the year covered in the 1934–36 survey *on the set of articles that this particular commodity represents in the index.*

These were *city expenditures* for food, housing, utilities, and miscellaneous and *regional* expenditures for clothing and house furnishings. Notice that *imputed* weights are now used in all groups, apparently for the first time in groups other than food.

The expression $p_a q_a$ is a pure definition, p_a being precisely defined above, but q_a now becomes a *hypothetical quantity* to represent a *set* of commodities.

Costs for each commodity for periods other than March 1935 were calculated by a linking procedure. Thus, for the June 1935 period ($k + 1$),

$$p_{k+1} q_a = \frac{p_{k+1}}{p_k} \cdot p_k q_a$$

and similarly for other periods Then aggregate costs for all commodities in a group are obtained by summation:

$$\Sigma p_k q_a; \qquad \Sigma p_{k+1} q_a, \text{ etc.}$$

for periods k, $k + 1$, etc.

Let $\Sigma p_0 q_a$ equal the average value of such aggregates for 1935–39. Then G_{0k} = index = $\Sigma p_k q_a / \Sigma p_0 q_a$ for varying values of k. A chain system will be required for the continuous index whenever the commodity list is changed.

2. All-Item City Indexes from City Group Indexes. Add the six group aggregates for a given city for any period k $(= \sum^{G} \sum p_k q_a)$. "This aggregate represents the cost *at a given date* of goods equivalent to those purchased by the employed wage-earners and clerical workers in a given city in 1934–36." [1] The base aggregate $\Sigma\Sigma p_0 q_a$

[1] See Bull 699, p 37 This bulletin is the source of all the technical information given here on the 1940 revision.

is the average of all such items as above in 1935–39. The index P_{0k} (city) $= \Sigma\Sigma p_k q_a / \Sigma\Sigma p_0 q_a$.

3. National Indexes (Both Group and All-Item). Cost figures for each city are weighted by population where each "city population" is the population of the metropolitan area of that particular city and of other cities in the same region and size class. For the new index of 1940, the population data of 1930 were used.

In current notation the above may be written, for any period k,

$$\Sigma p_k q_a(\text{U. S.}) = \sum^{\text{cities}} (\Sigma p_k q_a(\text{city}) \cdot w)$$

where $\Sigma w = 1$. The base of the index represents the averages of such costs for 1935–39, say $\Sigma p_0 q_a$(U. S.). Therefore, G_{0k}(U. S.) $= \Sigma p_k q_a(\text{U. S.}) / \Sigma p_0 q_a(\text{U. S.})$. The same procedure is followed for P_{0k}(U. S.) if all-item terms are involved.

A careful reading of the details of the above description will show that the technical methods of construction of this index have undergone almost no change since the 1935 revision. At the time of the latter, imputed weights had apparently been used only for the food index. In 1940 the system of imputation was extended[1] to the other groups wherever applicable. No changes occurred in the method of calculating all-item indexes or national indexes except for revision of population weights as later population data became available. Thus 1930 population figures were used in the 1940 revision, whereas the average of 1920 and 1930 populations was used when this method was first introduced in 1935. In the years since 1940, there have been further revisions of these population weights to keep them up to date with the indexes. Both 1940 census figures and 1942 estimates have been employed.

4. Wartime and Post-War Adjustments

One can well imagine that the U. S. Bureau of Labor Statistics thought, when the 1934–36 expenditure survey was completed and when plans underway for comprehensive revisions had been carried out, that it would be able to look forward to several years of a stable index of the fixed-weight aggregative form. But the revision was still in process when World War II broke out, and the United States was a belligerent almost before the new indexes were published. Another factor is important here: economic events do not wait for such things as formal declarations of war, and some of the wartime changes that spell havoc to fixed-weight index numbers were already taking place

[1] *Ibid.*, p. 32.

before Pearl Harbor. But, of course, they were all greatly accelerated after that date because a total war, of which World War II was the latest and greatest example, changes the character of total national production. It was not long after December 7, 1941, that many of the commodities that had first made their appearance in the consumers' price index in the 1940 revision began to disappear from the market for the very good reason that their production had either been greatly cut down or been stopped altogether in order to facilitate the national all-out effort to produce the tools and equipment of war. Automobile production of the private passenger car variety practically ceased, and private car operation was rationed into something that had only the slenderest relationship to those weights that had just been put into the index. By December 1942, changes in national consumption habits had become so great that the following list of commodities had been dropped:

automobiles
metal bed springs
sewing machines
silk stockings
radios
refrigerators
studio couches
vacuum cleaners
silk slips
washing machines
inner-spring mattresses
gas cook stoves
silk goods

Weights were increased for automobile repairs, street car and bus fares, and rayon yard goods and slips; they were decreased for gasoline, oil, and fuel oil. The food component of the index was revised in March 1943 to account for food shortages and rationing.[1]

These commodity disappearances are of significance for index number makers, because they show some of the changes in production and consumption to which an economy is exposed under the impact of total war, and they also reveal clearly the unreality of maintaining a constant commodity list and constant weights in an index number. If an index number is to measure price or production changes in a real world, these significantly real changes in goods produced or in quantities produced cannot be neglected. History has a way of finding out the deficiencies of a formal tool of this sort, as happened with this index during the war; for the changes were not enough, apparently, to satisfy some of the users and there were attacks upon the index's accuracy. But of that more in the next chapter.

[1] These details taken from "Description of the Cost of Living Index of the Bureau of Labor Statistics," revised May 1944.

5. Current Applicability of the 1934–36 Pattern

When World War II was over and peacetime production began to bring back into the markets some of the commodities that had disappeared during the war, the bureau returned to the use of the commodity lists and weights of the 1934–36 survey. The bureau claimed, no doubt with justification, that these lists and weights were more representative of the actual post-war economy than the lists and weights of the war period. The bulletin quoted last above goes on to say that the commodity list of the index "has been constantly adjusted since March 1935 *for necessary changes that have resulted from normal market developments* as well as from the rapid war time changes in consumers' goods available for civilians."[1] This position seems to indicate an abandonment of the policy of fixed weights for several years at a time in favor of one of change *whenever market developments justify*. If so, it abandons a policy that has held index number makers in its grasp practically since the publication of the Mitchell bulletin. Whether the 1934–36 family expenditure survey provides a pattern that is applicable in the late 1940's is a question of fact. It is one that deserves to be studied most assiduously, for the war brought many changes and there will be no more return to the prewar status in this postwar era than there was in the 1920's. The U. S. Bureau of Labor Statistics has recognized this situation clearly, for it began a postwar revision of its two most important indexes, the wholesale price index and the consumers' price index in the late 1940's.

At this point it should be recognized that the intention in Chapters 8 and 9 was not to write a complete history of these two indexes. Rather, the purpose was to obtain a skeleton picture of that history with emphasis upon the statement of, and any developments in, constructional methods, and by constructional methods is meant both the content and the form of index numbers that have received attention in these pages. These two chapters were written in order to provide a field for application of the principles laid down in the previous seven chapters, and the final chapter will be devoted to an appraisal of the result. The problem is two-sided: What has theory to say about practical index number construction? And can practical index number construction contribute anything to theory? The wartime history of these two index numbers provides an ideal chance to look into these matters.

[1] *Ibid.*, p. 4. Italics added.

CHAPTER 10

The Qualities of Two Famed Index Numbers, the Wholesale Price Index, and the Consumers' Price Index of the U. S. Bureau of Labor Statistics

10.1. THE ESSENTIALS OF GOOD MEASUREMENT RESTATED

The question to which this final chapter is devoted is a practical one. It is the very significant question, from the author's point of view, how the two index numbers whose history was sketched in the last two chapters measure up against the standards of good construction set forth in the previous seven chapters. Recommendations are implicit with respect to the defects shown in this comparison of practice with standards. For the purpose in hand it will therefore be well to restate the conclusions of this theory on how best to construct an index number of price or quantity change:

1. A field must be defined. The price and quantity content of the index will apply to this field only, in any exact sense. When it is desired to use the index as a measure of change covering a broader (or even a narrower) field, the question of the similarity of the two fields must be raised. With reasonable similarity of all the factors that enter into index numbers and into their ultimate source materials of price and quantity, the index may then be used as an *approximation*. If this similarity is not present, the index number should not be used to represent any field beyond its own.

2. Prices and quantities that enter into the index are, in prac-

tice, samples. The whole set of prices and quantities in any field is generally impossible to obtain for the construction of an index. The whole set must be represented, however. The proper procedure is to select a commodity list that will be representative in the sampling sense. In the list some commodities may represent only themselves whereas others may represent a group of other commodities not included in the index but of importance in the particular field of investigation. Thus all commodities in the field must be represented, either specifically or on a sampling basis by other included items that have a similar price history.

When a commodity represents others in this sampling sense, its weight must measure the importance of its entire group. In sound theory no exception to this rule can be contemplated. In what follows it is assumed that appropriate price and quantity data have been gathered.

3. The first step in calculation of an index involves the construction of a measure of change (price or quantity) *between two adjacent periods*. Measurement of this change guarantees the highest accuracy, because homogeneity as defined earlier is then at its maximum. For the maximum accuracy of measurement of this binary change, it is required that one of the formulas 3 to 6 be used, with preference given to formulas 3, 4, or 5 because each of these three can be stated in terms of the four aggregatives calculable from the prices and quantities of the two periods compared, Σp_0q_0, Σp_0q_1, Σp_1q_0, and Σp_1q_1.

4. Finally, the series comparison shall be constructed by chaining together the indexes of the several adjacent-period comparisons; for example, $P_{04} = P_{01} \cdot P_{12} \cdot P_{23} \cdot P_{34}$.

5. It has been contended throughout seven chapters that the above procedure results in the most accurate measurement of change between two adjacent periods and preserves all that is valid in a series comparison, namely, the direction of change. Although these are the most fundamental reasons for choice and are always to be preferred over any other the procedure here recommended has another important, though secondary, advantage, namely, that it greatly simplifies changes in the commodity list and in weights since every pair of years is treated independently of every other. These modifications can, therefore, be made whenever they occur in the actual markets, and the whole problem of recalculation of indexes when weights or commodity lists are altered is eliminated.

10.2. THE WHOLESALE PRICE INDEX IS GOOD BUT NOT GOOD ENOUGH

The U. S. Bureau of Labor Statistics wholesale price index, to begin with, is an index of wholesale and not of general prices. Despite the fact that it is frequently referred to as an index of general prices, there is no justification for such usage except to obtain the roughest kind of approximation. The reason is, of course, that the index leaves out so many of the prices in the price system, for example, wages, rents, security prices, and all retail prices. The fact that the index does not measure all prices is not a defect; it is only a property inherent in the commodity list that is priced. The famous Mitchell bulletin must bear a share of the responsibility for building up the wholesale price index as a general index, since it put so much emphasis upon a *general-purpose* index number It may, of course, be argued that "general purpose" does not imply "all prices," but at least there is a point here A wholesale price index has only one basic purpose and that is to measure changes in the level of wholesale prices, as defined by the makers, and of no other prices. For the index of the U. S. Bureau of Labor Statistics the commodity price coverage is probably one of its best features.

Not so much can be said for the formula used by this index. In its history, the index shows great progress but it falls just short of being the "best." It first used, as will be recalled, a simple average of relatives, and in this day everybody will agree with former Commissioner of Labor Statistics Ethelbert Stewart that "the less said about that index the better." [1] After Mitchell's recommendations the fixed-weight aggregative formula was introduced, and the first set of weights referred to the year 1909. The policy was established, either with the introduction of this formula or soon thereafter, of revising weights at intervals not too widely separated, for it was easily sensed that weights too far removed from particular prices were unsatisfactory. This bureau policy was well expressed in one of the bureau publications at the time of introducing 1919 weights: "In constructing the index numbers · · · in the 1921 report and in subsequent ones the prices were weighted by data from the 1919 census instead of the 1909 census data formerly employed. *This conforms to the plan contemplated by the Bureau at the inception of its weighted index number system in 1914 of revising the weighting factors every ten years as new census information should become available.*" [2] If the bureau's procedures have been rightly interpreted, the introduction of new weights, such as 1919

[1] Address published in *Monthly Labor Rev.*, December 1927.

[2] Bull. 367, pp 2–3 Italics added.

for 1909, was the occasion for a revision of the index numbers far into the past, and if this is so it raises a question of consistency to say the least. If 1909 weights are so far removed from 1923 prices as to be no longer satisfactory, then why upon the introduction of 1919 weights is it proper to construct indexes for 1909, 1910, and thereabouts with 1919 weights?

There is little doubt that the bureau's technical personnel did much serious thinking about this matter of weights during the 1920's for, in the address quoted above,[1] Commissioner Stewart announced what then must have seemed a new and radical policy on weights for the wholesale price index. It is indeed probable that these new ideas had grown from the discussions on index numbers that centered about the publication in 1923 of Fisher's *The Making of Index Numbers*, in which, of course, he advocated the Ideal formula in a binary comparison and that involved weights of base and given periods. Stewart reported the new policy in these words: "· · · the present plan is not only to add to the index new commodities as they appear but to re-weight with each succeeding census; that is to say, *the fixed weighting period has been entirely abandoned both in theory and in practice and the weights will be revised with each census period.*"[2] The new policy, unfortunately, did not turn out to be a permanent one, even though it was maintained until the revision of 1937, the details of which were given in Chapter 8.

A comprehensive revision of the wholesale price index was begun in 1949 It is to be hoped that progress in this revision, as in future ones, will be in the direction of changing weights as often as possible. How soon the theoretical ideal of using current weights only can be realized is a matter for conjecture. The changed weights at two-year intervals during the 1920's did not closely satisfy the requirements of our preferred formulas since, as Chapter 8 shows, the weights were generally several years older than the prices (for example, 1925–27 weights for the indexes of 1930–31). Nevertheless, the bureau procedures of these years furnish a precedent for developments in the right direction. The history of this period gives much evidence that index number formulas were undergoing searching examination and the appropriateness of current weights was being considered. The current index through the 1940's was constructed with weights belonging to the years 1929–30–31, and this incongruity offers one of the most cogent reasons for revision.

The bureau literature on the wholesale price index has described its

[1] See p. 124.
[2] See p 124 Italics added

formula as a chain index, in one article[1] remarking that the chain index method had been followed for twenty-nine years. The reference is to the year 1908, when the chain or linking procedure was first used in introducing substitute commodities into the index. The 1908 method involved obtaining prices for both the old and the new commodity for the period when the link was made. Then a single series of price relatives for this commodity and its substitute was calculated by linking the new series to the old. This is hardly a chain index in the sense in which the term is used in earlier pages, but it contains the germ of the idea. In a more complete sense the bureau did employ a chain index (as something more than just a single linked relative) from 1914 on,[2] whenever a change in the commodity list or a change in weights occurred. This makes the 1920's in all probability the best example of the bureau's application of the principle of the chain index. The practice indicates bureau familiarity with the principle and thus could facilitate the adoption of an out-and-out system of chaining individual index number links in accordance with the position which has been defended in these pages.

In brief, the wholesale price index offers a history of almost continual progress in construction procedures. The story is full of examples of experimentation with newer and newer ideas as these ideas have appeared in the general literature of the period. Witness the quotation describing the bureau's position in 1923 on ten-year revisions of the fixed weights and then, four years later, the authoritative statement by the Commissioner of Labor Statistics that fixed weights were being abandoned in favor of revision as rapidly as new weights were available. The most far-reaching proposals of this volume go very little beyond actual practices of the bureau in this period of its history in dealing with the wholesale price index.

10.3. THE CONSUMERS' PRICE INDEX IS ALSO GOOD BUT NOT GOOD ENOUGH

The consumers' price index is and always has been a large-city index. Even the national index is a large-city composite In no exact sense can it therefore be used to measure consumers' prices in small cities, in urban areas generally, in nonurban areas, or in particular industries. This is not a defect of the existing index but if these other indexes are needed in the interpretation of our national economy they must be specially constructed.

[1] Cutts and Dennis, in *J. Am. Statistical Assoc.*, December 1937.

[2] See corroboration by S R. Mitchell in *Monthly Labor Rev.*, August 1948.

The consumers' price index was established as a fixed-weight aggregative in the beginning, for it was not published on a continuous basis until several years after the publication of Mitchell's bulletin. Furthermore, the fixity of the weights has been a much more thoroughgoing part of this index than of the wholesale price index. This is doubtless due to the fact that in this index the weights have always been obtained from a comprehensive study of family expenditures, and such a study has to date been made only for the purpose of obtaining index number weights. No other purposes were served. In contrast, the weights for the wholesale price index come from the census and other periodical collections that were established with other uses in mind long before wholesale price index numbers had appeared. As has been said, the first expenditure survey was made in the years 1917–19, and the second in 1934–36. The first set of fixed weights were used from 1921 until 1940. The 1934–36 weights were introduced into a revised index in 1940, at which time World War II was upon us and in a few more months we were ourselves belligerents. Wartime changes in production and consumption, of course, brought tremendous distortions to the peacetime pattern of our economy, and many goods and services that had been available in time of peace soon disappeared completely. By the early 1940's the change in the consumption pattern of the typical American wage earner was so much in evidence that the bureau made wide adjustments in the commodity list of its consumers' price indexes and in weights. The index number makers were clearly conscious of the need for these adjustments in order that the index should remain a true measure of the market situation. The literature of these days provides much evidence that conflicting ideas were tugging for mastery in the minds of the technicians during World War II. Witness the frequent remark that the index represents "time to time changes *in the prices of a specific fixed market basket or shopping list of goods and services* bought by moderate-income urban (large-city) consumers,"[1] and match this against another emphasized statement that this market basket of goods "has been *constantly adjusted since March 1935 for necessary changes that have resulted from normal market developments* · · ·."[2] These seemingly contrary views need not convict the bureau of inconsistency. They point rather to the fact that the theory of a fixed-weight aggregative has been widely held and thoroughly accepted by so many index number students for so long that

[1] *The Consumers' Price Index,* Joint Committee on the Economic Report, 80th Congress, 2nd session, p. 3. Italics added.

[2] "Description of the Cost of Living Index of the Bureau of Labor Statistics," revised May 1944, p. 4. Italics added.

it recurs repeatedly in their thinking and remains as part of their practice. But the pressure of great changes in an economy also forces their attention in new directions, and the logical need for keeping an index close to the current market situation compels changes in the index when great economic changes occur in real life.

The wartime use of the consumers' price index as a basis for wage adjustments when prices were rising forced attention so strongly upon the need for constant revision of commodity lists and weights that it has become one of the most important of all existing influences, not only upon the demand for a revision of the consumers' price index, but also upon our whole thinking about this sort of measurement. The wartime adjustments, referred to above, came about under the influence of the most concentrated attack ever made on the accuracy of the index. This attack occurred during the war when the bureau's index reported a 23.4 per cent increase in consumers' prices between January 1941 and December 1942 and this finding was challenged by representatives of labor who, by a separate investigation, had measured the increase as 43.5 per cent. In these comments there is no need for, nor indeed any interest in, taking sides in the controversy between the bureau and labor representatives. What is important to recognize is that the labor people made certain charges about the accuracy of the index and that these charges at once went to the very heart of the problem that has been central in this book, that is, the question of currency of commodity lists and weights Briefly, labor charged (1) that prices important in the workingman's budget but not in the index had risen higher than the prices which were included in the index (this, if true, meant bad sample representation); (2) that shifts in the character of food consumption were not reflected in food weights; (3) that the index did not reflect the quality deterioration which had occurred in the markets; and, finally, (4) that the index for thirty-four cities did not properly represent all urban areas and that the nonrepresented areas were among the ones most influenced by wartime changes.

The fourth point above is correctly taken but is no proper criticism of the existing national index. If an index for small cities is needed, it must be constructed anew. The other three points all bear upon the question of the validity of the commodity lists and weights used at that time, and they were in reality a devastating attack upon the whole theory of a fixed-weight aggregative. Here, as so often happens, it took the radical economic changes of a period of great national emergency to force the technicians to consider the importance of a change in method that may be important in normal peacetime as well as in wartime. The difference between wartime and peacetime is one of

degree. But the important principle that is involved deserves a final and forceful statement: the only measure of the importance of p_0 is q_0, and the corresponding weight of p_1 is q_1.[1] The quantity q_a has nothing to do with the situation. Out of these facts were evolved the "best" formulas of this book. And it deserves to be emphasized that this book has not thereby propounded something new and original. All these formulas are fifty or seventy-five or more years old. The chain system of combining them into a series is two-thirds of a century old, and its origin belongs to Alfred Marshall. The trouble is that we have forgotten the advice of the sages of long ago, and it has taken the economic distortions of a total war to show up the most basic weaknesses in our index number constructions.

This chapter and the book close on a hopeful note. As a result of labor's criticism of the consumers' price index a committee of experts representing the American Statistical Association was asked to investigate the charges and to make recommendations. These recommendations represent a landmark in the history of index numbers. Here in brief are the basic ones, and they include those with which this book has dealt:

1. Provide separate regional indexes.
2. Provide indexes for small cities.
3. Provide indexes for nonfarm communities in the United States.
4. Prepare and publish indexes immediately for:
   (a) Communities of different sizes.
   (b) Specialized industrial areas, for example, coal mining, textiles.
5. Develop indexes of area differences.
   (The first five recommendations involve the construction of new indexes.)
6. Make frequent small sample studies of family expenditures and income and make a comprehensive study every five years.
7. Make special studies of changes in the specific goods and services that the bureau prices.

After the last two it is but a short step to demand that quantity data be gathered annually the same as price data. Indeed, for our most important quantity index, the price weights could now be made available every year.

These changes are what index number practice has been in need of for years. And they apply not only to consumers' price indexes. They apply to all.

[1] See the development in Chapter 4.

BIBLIOGRAPHICAL REFERENCES CITED IN TEXT

I. BOOKS AND PERIODICAL ARTICLES

1. Jesse M. Cutts and Samuel J. Dennis: "Revised Method of Calculation of the Wholesale Price Index of the United States Bureau of Labor Statistics," *Journal American Statistical Association*, Vol. 32, pp. 663–674, December 1937.
2. F. Y. Edgeworth: *Papers Relating to Political Economy*, Vol. 1, Papers H and I, Macmillan & Co., Ltd., London, 1925.
3. Irving Fisher: *The Purchasing Power of Money*, The Macmillan Co., New York, 1911.
4. Irving Fisher: *The Making of Index Numbers*, Houghton Mifflin Co., New York, 1923.
5. Irving Fisher: "The Total Value Criterion: A New Principle in Index Number Construction," *Journal American Statistical Association*, Vol. 22, pp. 419–441, December 1927.
6. Margaret Hogg: "A Distortion in the Cost of Living Index," *Journal American Statistical Association*, Vol. 26, pp. 52–57, March 1931.
7. W. Stanley Jevons: "A Serious Fall in the Value of Gold Ascertained," 1863; reprinted in *Investigations in Currency and Finance*, Macmillan & Co., Ltd., London, 1884; reprinted 1909.
8. A. A. Konüs: "The Problem of the True Index of the Cost of Living," *Econometrica*, Vol. 7, pp. 10–29, January 1939.
9. Alfred Marshall: "Remedies for Fluctuations of General Prices," *Contemporary Review*, March 1887.
10. Wesley Clair Mitchell: *The Making and Using of Index Numbers*, Part I of *Index Numbers of Wholesale Prices in the United States and Foreign Countries*, United States Bureau of Labor Statistics, *Wholesale Prices*, Bulletin 173, 1915, republished as Bulletin 284, 1921. Part I reprinted as Bulletin 656, 1938. Referred to in footnotes as Mitchell, Bull 284.
11. Bruce D. Mudgett: "The Cost of Living Index and Konüs' Condition," *Econometrica*, Vol. 13, pp. 171–181, April 1945.
12. A. C. Pigou: *Wealth and Welfare*, Macmillan & Co., Ltd., London, 1911.
13. A. C. Pigou: *Economics of Welfare*, Macmillan & Co., Ltd., London, 1920.
14. Henry Schultz: "A Misunderstanding in Index Number Theory: The True Konüs Condition on Cost of Living Index Numbers and Its Limitations," *Econometrica*, Vol. 7, pp. 1–9, January 1939.
15. Ethelbert Stewart: "Explanation of the New Indexes of Wholesale Prices of the United States Bureau of Labor Statistics," *Monthly Labor Review* of United States Bureau of Labor Statistics, December 1927.
16. Correa Moylan Walsh: *The Measurement of General Exchange Value*, The Macmillan Co., New York, 1901.
17. Correa Moylan Walsh: *The Problem of Estimation*, P. S. King and Co., London, 1921.
18. Allyn A. Young: "The Measurement of Changes in the General Price Level," *Quarterly Journal of Economics*, Vol. 35, pp. 557–573, August 1921.

II. GOVERNMENT REPORTS AND DOCUMENTS

A. Relating to Wholesale Prices

1. Committee of Finance of the United States Senate: *Wholesale Prices, Wages and Transportation,* Report No. 1394, 52nd Congress, 2nd Session, March 3, 1893.
2. United States Department of Labor: Bulletin 27, 1900; Bulletin 39, 1902.
3. United States Bureau of Labor Statistics: Bulletins 173 and 284 (the Mitchell study, listed as No. 10 above); Bulletins 181, 269, 296, 320, 367, 453, 472, 473, 493, 521, 543, 572.
4. S. R. Mitchell: "Wholesale Price Index: Policy on Revisions and Corrections," *Monthly Labor Review,* August 1948.

B. Relating to Retail Prices and Cost of Living

1. United States Commissioner of Labor: 6th annual report, 1890, expenditure study of 3260 families; 7th annual report, 1891, expenditure study of 5284 families; 18th annual report, 1901–02, expenditure study of 25,440 families.
2. United States Bureau of Labor: Bulletin 54, September 1904, first published index of food prices.
3. United States Bureau of Labor Statistics: Bulletins 357, 396, 638, 699; "Standard of Living of Employees of Ford Motor Company, Detroit," *Monthly Labor Review,* June 1930, pp. 11–54; *Monthly Labor Review,* September 1935; "Description of the Cost of Living Index of the Bureau of Labor Statistics," revised May 1944; *Monthly Labor Review,* August 1948.
4. *The Consumer Price Index,* Joint Committee on the Economic Report, 80th Congress, 2nd Session, 1949.

Index